KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

TRADING IN TRANSPARENT AND OPAQUE FINANCIAL MARKETS

Dissertation presented
to obtain the
degree of Doctor in
Business Economics

by

Geofrey Tombeur

Number 501

2015

Daar de proefschriften in de reeks van de Faculteit Economische en Toegepaste Economische Wetenschappen het persoonlijk werk zijn van hun auteurs, zijn alleen deze laatsten daarvoor verantwoordelijk.

Doctoral Committee

Advisors

Prof. dr. Gunther Wuyts	KU Leuven	advisor
Prof. dr. Hans Degryse	KU Leuven	co-advisor

Members

Prof. dr. Carole Gresse	Université Paris-Dauphine
Prof. dr. Vincent van Kervel	Pontificia Universidad Católica de Chile
Prof. dr. Kristien Smedts	KU Leuven
Prof. dr. Rosanne Vanpée	KU Leuven

Chairperson

Prof. dr. Liesbeth Bruynseels	KU Leuven

Acknowledgments

On a day in May 2010 I was studying in the central library of the KU Leuven, as I had done so often during the past years. I had just submitted my master thesis, and so had all the students that I had guided during my past year as a teaching assistant. It was at that moment that I decided, after a long period of hesitation, that I wanted to apply for a Ph.D. position. More than five years later this adventure has now almost come to an end. Looking back at this period, I think of it as a great time, during which I met some extraordinary people and learned so much new things. Also in my personal life there have been quite some changes since then. Most importantly, I am now the father of two amazing daughters. Finally coming to the completion of the Ph.D. process is an achievement that I could not have accomplished on my own. Therefore I would like to take the opportunity to thank all the people who have supported me, in one way or another, throughout the process.

First and foremost, I would like to express my gratitude to my advisors Gunther Wuyts and Hans Degryse. Gunther, during the period that I was working as a teaching assistant, you suggested me to apply for the vacant Ph.D. position. I remember it took you some time to convince me that I would be suited for an academic research position, but I have not regretted my decision ever since (or at least most of the time I haven't). Thank you for offering me this incredible opportunity. Hans, you became my co-advisor after two years, and you gave me the opportunity to work on an incredible dataset. This has been very challenging (and at times frustrating), especially in the beginning. But it allowed me also to conduct relevant empirical research, and I learned a lot of new skills! During the process you were both always available for feedback. You tried to answer the numerous questions that I sometimes had, and you encouraged me to follow through when results were not as expected, or at those times the workload seemed unmanageable. You also stimulated me to submit our work to conferences, which helped me to push some boundaries and improve our papers. As a result I was able to participate in various interesting international conferences in Rotterdam, Parma, Milan, Stuttgart, Buenos Aires and finally the EFA Annual Meeting in Vienna, all of which have been exciting experiences for me. On top of that you were also willing to present our work on seminars and conferences. Now, with the completion of my Ph.D. in sight, you have supported me in the search for a new job. Also personally I got to know you better, and I always much enjoyed our talks. So thank you for all the years of pleasant collaboration.

I am also grateful for having such an inspiring group of committee members. I would like to thank Carole Gresse, Vincent van Kervel, Kristien Smedts and Rosanne Vanpée for all the time they invested in helping me to improve this dissertation. Carole, you discussed one of the earliest drafts of the paper on opaque trading. Your comments helped to improve the paper a lot, and have definitely contributed to get the paper accepted at so many conferences this year. Vincent, you provided me with the dataset for the second and third paper, downloaded manually. I can only imagine what a painstaking task this must have been. You helped me to get started with the data by providing me with some of your code, and afterwards you almost instantaneously helped me whenever I had a question. So I think it's fairly safe to state that without you this dissertation would look very different. Kristien, I took my first steps into academic research under your wings, as you were the supervisor of my first master thesis. Triggered by this experience I became a teaching assistant in the following year. For the last five years it was my pleasure to have you as a neighbor in the corridor. Rosanne, you were always among the first persons to read early drafts of my papers. Thank you for always being available. Finally, Liesbeth, I would also like to thank you for all the nice coffee talks.

Although performing academic research is quite solitary work most of the time, daily life at the faculty was far from lonesome. I would like to thank therefore the amazing group of Ph.D. students at the AFI department for all their friendship and support. Commuting to Leuven almost every day for five years has been an investment of more than two hours per day, but thanks to the positive atmosphere it was all worth it. Some persons I was able to work with more closely. Christophe, Elisabeth, Ines, Jorden, Jorgen, José, Karen 1, Karen 2, Katrien B, Lisa, Nathalie, Nima, Randy, Rosy, Sander DG and Sharon, I really enjoyed working with you guys. Unfortunately I did not have the opportunity to get to know everybody of the AFI group as well as I would have wanted to, but every time we had lunch together, sat on the train, or just talked in the corridor, it was always a pleasure. So thank you Anastasios, Ann-Florence, Aurélie, Ben, Dieter, Els, Evelien, James, Jan-Francies, Jeroen, Kathleen, Katrien C, Kristof, Lore, Rianne, Robin, Roel, Sanja and Sofie for being a part of the awesome AFI group.

I'd also like to express my gratitude to some colleagues in particular. Katlijn, as a Ph.D. student I think you have many qualities, in particular you are a very meticulous worker and ambitious researcher. As a person, you have even better qualities. Thank you for always being such an empathic, thoughtful and helpful colleague. Tom, although we studied business engineering together for five years, I had never met you before we started the Ph.D. adventure in October 2010. This is remarkable given your generous laugh which has become world famous in our faculty. It makes me believe that you enjoyed your time here at least as much as I enjoyed your weekly updates on KFC Diest. Thank you for being a wonderful person. There are a lot of persons that can be thought of as having a very special opinion about life, but Wouter, I think you top them all. Your most surprising moment, however, was the day you bought me chocolates. And for that, I want to thank you. Weidong, you also have a very surprising side. I always thought of you as one of my cleverest colleagues who also knows how to enjoy life. But in Argentina I discovered you are most of all a tango dancer disguised as a Ph.D. in finance. I was very happy that after two years of working finally another Ph.D. student joined the group with the same research interests. Matthias, I appreciate all our microstructure talks and your advice on my research. Most of all, thanks for letting me co-author an opinion in De Standaard on HFTs – your topic. Reina, of all my colleagues I have known you the longest, and I am sure we will keep in contact in the future, as you will continue to invite yourself to my house. Thank you for arranging numerous lunches, dinners and barbecues. Mathijs, you are one of the most spontaneous and self-confident persons I have ever met. Despite of that your are also one of the most likable persons. You have an amazing knowledge of useless facts and I am ever grateful that you cared to share these with me every Monday morning when you passed by my office. Too bad only half of it was actually true, but still it always succeeded in cheering me up. Your coffee has also provided the necessary fuel for this dissertation. Just keep on sending those memes and inviting me to the Oude Markt. I promise I will join you one day! Marjan, you are one of the brightest and hard working Ph.D. students of our group, and in addition you succeed in being the heart of the social network. You are always there for other colleagues, including myself, whenever they need advice, or just want someone to listen. Thank you for being such a remarkably sympathetic person. You have become a true friend (as is evidenced by choosing my Ph.D. defense over Matt Berninger) and I hope we will keep on seeing each other in the future, when you have a flourishing academic career.

Finally, I want to thank three colleagues with whom I shared an office. Stien, you were my first office mate when I started as a teaching assistant. From the very first moment we met you made me feel welcome, and this has surely contributed to my decision to apply for the Ph.D. position. You shared a lot of tips and tricks that taught me how to give proper guidance to master thesis students. But most of all I enjoyed our talks about everyday life. Although we don't often see each other now, every time we do meet it's just like we pick up the conversation where we left off the last time. Sander, we started this adventure together. I remember our first day in the office as if it was yesterday. You were my partner for some courses in the doctoral program, and we often read and commented on each other's early work, but most of all you were my buddy. I appreciated our conversations about the little things in life, but also a lot of serious stuff, during our daily walks between the railway station and the faculty. Thank you for being my friend. Yannick, for five years we sat opposite to each other on almost every weekday (with some computer screens in between though). We shared important moments in our life, and by now, I think you are one of the people who knows me the best. Thanks for the numerous times you gave me advice (research-wise, but mostly non-research-wise), for motivating me whenever I had bad days, for all the laughs we shared, and for the great time in Buenos Aires. I am really fortunate to have had such a thoughtful and enjoyable friend as my closest colleague. Our daily talks is what I will miss the most in the post-Ph.D. life.

Next to my colleagues I could always count on many other friends to support me. Although I don't think any of you guys really know what I have been doing for all these years, besides that I studied something about exchanges and an elusive concept called 'dark pools'. (I certainly hope I could have enlightened you somewhat during my presentation.) Thanks for encouraging me, and putting things in perspective from time to time. During evenings in 'Komaf' and 'De Wipschutter', quiz nights or on weekends and holidays you provided the necessary distraction that I needed from time to time. Even though sometimes it was hard to find the right balance, and the Ph.D. work made me miss out on some things, such as the day that I was busy preparing a seminar while we were on a holiday in Turkey. While a certain quote that was printed on the t-shirt designed for my bachelor party suggested something else, you really are the most amazing group of friends.

Special thanks also go to Vicky, who helped me countless times during my business engineering and financial economics studies. For the latter program we wrote our master thesis together, and the first chapter of this dissertation is a direct follow-up of that research. My second home during the exam periods was the central library of KU Leuven. Leen, during these periods I could always count on you for distraction. Looking back at that time makes me truly nostalgic.

I also want to express my gratitude to my family. I am blessed with many siblings: two brothers, three sisters, two stepbrothers and a stepsister. Although you guys considered me to be somewhat aloof and smug from time to time (and I am currently enjoying that you have to look up what these words actually mean) you have been very supportive, not only during the past five years, but our entire lives. The last two years you have all been very eager to babysit on one of my daughters, which was a big help when the workload was soaring.

My sincerest appreciation also goes out to my family-in-law. Bart and Chris, during the ten years that I have known you, I must have seemed like an eternal student. Nevertheless you supported me the whole time. Today my student years have finally come to an end. Thank you also for taking care of Juliet every Thursday. I know you look forward to it as much as she enjoys it.

I would never have come to this point without the lifelong support of my parents and stepparents. Sunday family meetings have been the necessary moments of rest during the sometimes stressed weeks. I am forever grateful for all the chances you have given me, and your encouragements during all my endeavors. You raised me to the person I am today, and I am sure that you are proud to have a third generation doctor (though of Philosophy instead of Medicine) as a son. Thank you for the homegrown vegetables, kitchen table discussions, hotdog Wednesday afternoons, holidays by the seaside, ... and much more. But mostly, I would like to thank you for the warm home.

Finally, I would like to thank Joke, my wife. We are together for ten years now, and half of that time I have spent concurrently working on this dissertation. I remember that when I told you I would be pursuing a Ph.D. you had mixed feelings about it: you were proud, but at the same time you knew it meant sacrificing a lot of free time. I know these past five years have not always been easy for you. They included both moments of intense happiness and hard times. Nevertheless, you always stood by my side. You took care of me, and all the household work, when I did not have time to do so. Thank you for being so understanding. As I have evolved as a person, we have evolved as a couple, you became my wife, we bought a house and you gave birth to two beautiful daughters, Juliet and Olivia. Their birth has been the most significant change in my life, and living everyday life with these wonderful little persons has really helped to put things in perspective. The three of you are truly an inspiration. For that, and all those other moments during the past ten years, thank you.

Geoffrey Tombeur
Leuven, November 2015

Conferences

This dissertation benefited from numerous comments and suggestions received at several international conferences and seminars. The first chapter was presented at the 2011 3L Workshop (February 2011, Brussels, Belgium), the doctoral symposium of the Fourth Erasmus Liquidity Conference (July 2011, Rotterdam, Netherlands) and a KU Leuven seminar.

The third chapter was presented at the 2014 3L Workshop (April 2014, Brussels, Belgium), the fourteenth Belgian Financial Research Forum (May 2014, Louvain-la-Neuve, Belgium), the XVI Workshop on Quantitative Finance (January 2015, Parma, Italy), The Developments in Securities Markets: Trends, Risks and Policies conference organized by Bocconi Univeristy and CONSOB (February 2015, Milan, Italy), the Doctoral Consortium of the third European Retail Investment Conference (ERIC) (April 2015, Stuttgart, Germany), the sixth World Finance Conference (July 2015, Buenos Aires, Argentina), the 42nd European Finance Association (EFA) Annual Meeting (August 2015, Vienna, Austria), the 11th Annual Central Bank Workshop on the Microstructure of Financial Markets (October 2015, Dublin, Ireland), the Inquire 25th Anniversary Autumn Seminar (October 2015, Athens, Greece) and seminars at KU Leuven and Manchester Business School. This chapter received the doctoral student best paper award on the third European Retail Investment Conference (ERIC).

Contents

List of Figures

List of Tables

General Introduction

This dissertation contributes to a growing literature that analyzes the way financial markets are structured and the trading process itself. Ever since the inception of organized financial markets, their design has been in constant evolution, adapting to new technologies and regulation in order to provide investors with the most efficient and effective way to trade securities. Especially the last decades have witnessed some dramatic changes owing to advances in computing power and communications technology, and further stimulated by a transformation in the regulatory framework. This dissertation touches upon three of these recent changes: increased transparency, market fragmentation and the rise of opaque trading mechanisms.

Most financial securities, especially equities, are now traded on electronic platforms through transparent limit order books. As a result, overall transparency in financial markets increased steadily. The move to electronic trading has also been accompanied by a fragmentation of the market. Whereas traders used to be constrained to trading on the listing exchange, they can now trade on a multitude of trading venues. Furthermore, following the greater transparency of today's markets, several tools have been developed that allow traders to hide their intentions by trading opaquely. Traders can do so either by using specific order types, labeled *hidden orders*, or by trading on completely dark venues, so-called *dark pools*. This dissertation contributes to the literature on transparent limit order markets in Chapter 1, by investigating why short-term returns can be predicted using public order book information. The analysis is extended in Chapter 2, which examines whether information from other trading venues can be used on top of information from the own venue to predict returns, thereby contributing to the literature on market fragmentation. Chapter 3 aims to bridge two related streams in the literature by studying the interaction between the two distinct ways to trade opaquely: trading with hidden orders on transparent trading venues and trading in completely dark trading venues.

The remainder of this general introduction is structured in two parts. First, a brief description of the trading landscape and different types of traders is meant to be a general introduction into the world of securities trading. Second, a chapter overview discusses the specific contributions of this dissertation.

Traders and the Trading Landscape

In this research market variables are studied that are affected by the trading process and trader behavior. This section therefore provides an introductory overview on financial market organizations and the different trader types that populate the market. The most general distinction in market designs is between quote-driven markets and order-driven markets. In quote-driven or *dealer markets*, traders have to trade via one or multiple intermediaries, also called dealers. These dealers quote ask prices at which traders can buy a certain amount of the security (the depth), and bid prices at which they can sell. In this type of market liquidity is thus determined by the willingness of dealers to transact.

Many financial markets, however, are now organized as order-driven markets. In such markets there are no designated intermediaries, instead traders interact directly with each other. For instance, most stocks have a continuous trading session throughout the trading day that is organized as a so-called *limit order market*. Traders in these markets have the choice between two basic order types: market orders and limit orders. A market order to buy (sell) a security is similar to what is available in a dealer market: it immediately executes against the best outstanding orders on the ask (bid) side of the limit order book. A limit order is a priced order that can be either marketable or non-marketable. A marketable limit order has a limit price that is set such that it immediately executes, and thus behaves as a market order. While a non-marketable limit order does not immediately execute upon its submission, but instead is added to the already outstanding orders in the book. These limit orders can then get executed later (against future market orders) according to a set of priority rules. Usually price and time priority is enforced, meaning that first orders at better prices get executed, and at the same price those orders that were submitted first also get executed first. In this type of market liquidity is determined by the willingness of traders to submit limit orders. The limit order book is the mechanism that is central to the trading process in all chapters in this dissertation.

Besides limit order books other types of order-driven markets also play a role in today's trading landscape. Some stocks are traded in periodic auction markets, and on a number of exchanges there is also an opening and closing auction to determine the opening and closing price, respectively. In addition to being traded on the main market, some stocks are also traded in satellite markets that employ a crossing network mechanism. This means that buy and sell orders are matched as much as possible (periodic or continuously), but the price at which they execute is derived from the main market. Finally, orders can also be internalized by broker/dealers when the latter trade against orders of their clients in-house.

An important aspect of market organization is also whether the market is concentrated or fragmented. In a *concentrated* market there is only a single trading mechanism where all potential buyers and sellers meet. In a *fragmented* market there are multiple trading mechanisms by which transactions can be executed, possibly catering to different traders. For instance, institutional traders generally have much larger order sizes compared to retail traders. As such there have always existed parallel trading mechanisms to facilitate these institutional transactions, such as the upstairs market where brokers used to arrange these so-called block trades. When trading moved to an electronic platform, facilitating block trades has been the driver behind the inception of dark pools. Of course, markets can also be fragmented when trading venues with a similar design compete with each other, as is the case when multiple limit order books compete for order flow. While Chapter 1 focuses on a concentrated market, Chapters 2 and 3 describe such a fragmented market setting.

Transparency is another feature of market design. We distinguish between two types of transparency: pre-trade transparency and post-trade transparency. *Pre-trade transparency* generally refers to the extent of information that is displayed to market participants about the available liquidity against which can be traded. For instance prices and depths, or the identity and behavior of the traders who have submitted orders. In limit order markets traders can generally observe price and depth information to a great deal (while traders remain anonymous), as it is a requirement for the proper functioning of such a market. It is also mandated to a certain extent by regulators. Other trading venues, such as dark pools, are non-transparent or opaque by design. The display of large orders would alert the market and can lead to a hike in trading costs for these orders. *Post-trade transparency* refers to the reporting of transactions to market participants. For many markets, transaction reporting is mandatory in real-time, provided some exceptions. In this dissertation, transparency generally refers to pre-trade transparency.

All of the chapters describe, at some point, the behavior of traders, for which we rely on theory models. In these models traders are classified into different categories along a number of dimensions. Most importantly, traders can be classified according to their motivation. The two main trader types are informed traders and liquidity traders. Informed traders have private information about the value of a security that other traders do not possess. As they want to exploit their information advantage by trading on it, prices become informative. Liquidity traders trade for non-informational reasons, e.g., because they experience shocks to their endowment. Sometimes traders with other motivations are also distinguished, such as predatory or parasitic traders, who want to exploit the knowledge that they have about the orders of other trades, e.g., by front-running large orders.

Another distinction between traders is from the point of view of a transaction. Each transaction has an 'active' side and a 'passive' side. The active side has initiated the trade by demanding/taking liquidity with a market(able) order, while the passive side has facilitated the trade by supplying/providing liquidity with a limit order. When both orders are matched a trade gets executed. In limit order markets, whether a trader is a liquidity demander or supplier is endogenously determined, as traders can choose between different order types. Professional liquidity suppliers are referred to as market makers. Other dimensions according to which traders are classified include size (large traders versus small traders), speed (fast traders versus slow traders) and trading technology (human traders versus algorithmic traders). High-frequency traders (HFTs) are both fast and algorithmic in nature. Finally, both in theory and in practice it can be important to distinguish between retail traders and institutional traders. Retail traders are characterized by a smaller transaction size, a lower level of sophistication and are generally perceived to be 'uninformed', while institutional traders can have large (parent) order sizes and may be more sophisticated.

Chapter Overview

Over time financial markets have become increasingly transparent. The amount of information that is disseminated to market participants in real-time is now greater than ever. The sale of market data and direct access feeds is also a growing source of revenues for trading venues (Cespa and Foucault, 2014). Market data are the primary input for many trading algorithms. Information from trading venues is thus likely to be relevant for many traders. Previous research shows that one way in which information from the limit order book can be used is to predict very short-term returns (see e.g., Harris and Panchapagesan, 2005; Cao, Hansch, and Wang, 2009; Brogaard, Hendershott, and Riordan, 2014). In Chapter 1 of this dissertation we show for a sample of 88 stocks listed on the Spanish stock market in 2003 that we can predict intraday returns using *imbalances* between the bid and ask side of the limit order book. A more liquid bid than ask side predicts positive returns, and vice versa. Since this is a robust finding across markets and over time, we empirically examine what drives return predictability. We distinguish between two potential explanations: (1) the order choice hypothesis, and (2) the informed trading hypothesis. According to the first explanation traders adapt their order choice on the basis of the state of the order book, i.e. they become more (less) aggressive when the book on their side of the market is thicker (thinner) (see e.g., Parlour, 1998; Ranaldo, 2004). This causes transitory price effects, and thus returns are predictable over a short term. The second explanation entails that informed traders submit orders to the limit order book and naturally cluster on one side of the market, which causes the book to become informative (see e.g., Bloomfield, O'Hara, and Saar, 2005; Kaniel and Liu, 2006). This information is then later on incorporated into prices. Our evidence is more in line with the former explanation. First, we find that the bulk of return predictability is transitory, as longer horizon returns are less predictable.

Second, we do not find a clear intraday pattern as we would expect when predictability is driven by informed limit orders. Finally, cross-sectional analysis reveals that predictability is negatively related to informed trading, and positively to the depth of the order book. The latter is in line with the order choice hypothesis, as queue-jumping activity (switching in the order choice) is more likely in deeper, more competitive order books.

Chapter 2 extends this analysis to a *fragmented* market. In a fragmented market there is not a single venue where all trading activity concentrates. Instead, multiple trading venues compete for order flow. This is exactly the type of setting in which trading currently takes place in both the US and European markets. In the US, Regulation National Market System (Reg NMS) allows for competition between exchanges, Electronic Communications Networks (ECNs) and dark pools. In Europe, the Markets in Financial Instruments Directive (MiFID) has abolished old member state concentration rules, and encouraged competition between trading venues. Newly established Multilateral Trading Facilities (MTFs), such as Bats Chi-X and Turquoise, are now competing with traditional Regulated Markets (RMs). In addition, orders can be internalized by Systemaic Internalizers (SIs) and a large part of volume is executed in the unregulated Over-the-Counter (OTC) market. See, e.g., Degryse, de Jong, and van Kervel (2015) and Gresse (2015) for implications of fragmentation for market quality in Europe. Our focus here is only on the so-called *lit* market, in which trading venues (RMs and MTFs) with transparent limit order books compete for order flow. Using a sample of 30 stocks listed on Euronext in 2009-2010 we show that on each of the lit trading venues where the stocks are traded (Euronext, Chi-X, Turquoise and Bats), as well as on the consolidated market (taking all the order books together) returns are predictable with imbalances obtained from the *own* venue. On top of that, imbalances obtained from *other* trading venues also contribute to return predictability on the own venue, the more so when these imbalances are recorded at more competitive prices. We also show that prices from different trading venues tend to adjust to each other. These prices are not determined in isolation of each other, and are bound by no-arbitrage relationships as they are for the same security. Euronext, which is the listing exchange, dominates in price discovery as returns on Euronext are least predictable by the adjustment measure that we use. By contrast, returns on Turquoise and Bats are most predictable. Lastly, we show that imbalances on a trading venue can better predict returns on another trading venue when it has a relatively larger market share, at least when the former venue is an MTF and the latter is the listing exchange. As MTFs attract a larger fraction of trading activity they become more relevant in terms of order choice considerations. This better integration in the market causes their order book to become a better predictor of returns.

In Chapter 3 we broaden our view to the entire market, including also *dark* trading venues, and mechanisms to trade *opaquely* in general. In contrast to the limit order books studied in Chapters 1 and 2, dark trading venues are (by definition) non-transparent. This means that traders submit orders to these venues without being able to properly assess the liquidity that is available, i.e. there is a large uncertainty about the execution probability of their order (for an overview of the literature on dark trading, see Degryse, Tombeur, Van Achter, and Wuyts, 2013). In addition, lit trading venues offer traders with the possibility to submit *hidden* orders, and by doing so create a pool of opaque liquidity that is embedded in the lit market (De Winne and D'Hondt, 2007; Bessembinder, Panayides, and Venkataraman, 2009). One of the primary advantages of these opaque trading mechanisms is that they allow traders to hide their trading intentions. As overall transparency of financial markets has grown with the increased reliance on electronic order books so have exposure costs. These are costs associated with the display of orders in financial markets and they can arise for various reasons. In general, these arise because other traders (who are not a natural counterparty for the trader) observe the displayed order and react to it in an unfavorable way (Brunnermeier and Pedersen, 2005; Buti and Rindi, 2013). Managing exposure costs is thus key to reducing overall trading costs. Using a sample of 27 Dutch index stocks from November 2007 until September 2010 we investigate whether both types of opaque trading (trading with hidden orders on otherwise transparent venues versus trading in completely dark venues) are complements or substitutes to each other. We find that they are substitutes, but dark trading is a better substitute for hidden order trading than the other way around. Both types of opaque trading are also driven by different market conditions. Hidden order trading is relatively more important when markets show a greater volume, are less deep and have a lower spread, and when less traders use Smart Order Routing (SOR) technology, which is a type of algorithm used to trade automatically on different trading venues. Dark trading is relatively more important when markets have lower overall volume and more traders rely on SORs. Both types of opaque trading are negatively affected by algorithmic trading.

Finally, it should be noted that we aim to make each chapter readable as a stand-alone paper. This approach implies some repetition between chapters, however. In particular, Chapter 1 and Chapter 2 use roughly the same methodology to predict short-term returns, and part of the related literature is also similar. Chapter 2 and Chapter 3 use a comparable dataset, and thus the institutional details are the same.

Chapter 1

Limit Order Book Information and Return Predictability

1.1 Introduction

Order-driven markets, governed by public and electronic limit order books, are by now widely used mechanisms to transact financial securities, particularly so for equities (Jain, 2005). Through the limit order book (henceforth also referred to as 'the order book' or 'the book') market participants can submit orders and are simultaneously updated about the liquidity available in the market. Traders can generally observe in real-time bid and ask prices and their associated depths up to a certain level in the book. Previous research shows that short-term intraday returns are predictable using observable information from the state of the order book in a number of markets: the New York Stock Exchange (Harris and Panchapagesan, 2005), the Australian Stock Exchange (Cao, Hansch, and Wang, 2009) and the Tokyo Stock Exchange (Jain, Jain, and McInish, 2011; Yamamoto, 2012).

Our goal is to examine what actually drives this type of return predictability. More specifically, we distinguish between two potential explanations. On the one hand, when traders use the state of the book to condition their order submission strategy on, the book may generate predictable patterns in order flow. This could in turn cause predictable, but short-term, price effects. On the other hand, if informed traders submit limit orders to the book they can make the book informative on the fundamental value of a security, thereby causing returns to be predictable. In the latter case order book information should also have a permanent impact on prices.

We obtain data from the Spanish Stock Market Interconnection System (SIBE) for the full calendar year 2003. These stocks are traded continuously on a limit order book platform without any designated market makers, which makes the market structure purely order-driven. Moreover, during this period there was no competition from alternative trading venues, so this environment allows a clean analysis to test limit order book theories. Similar data are used by Pascual and Veredas (2009a), Pascual and Veredas (2009b), Theissen and Wuyts (2010) and Pardo and Pascual (2012).

To examine the drivers of predictability we first verify its existence. To do so we take snapshots of the limit order book every 5 minutes. For each snapshot, we compute three types of measures capturing information about different levels of ask and bid prices, and depth offered at each price. Our first measure is the $Slope$ of the limit order book, as defined in Naes and Skjeltorp (2006), Intuitively, this measure combines price and depth information into one number and can be interpreted as an elasticity. Second, we use the $Depth\,(X)$ measure from Degryse, de Jong, and van Kervel (2015), which captures depth available in the limit order book within a certain interval of X basis points around the midquote between best ask and bid.[1] Third, we apply the cost of a buy and sell trade, CT, of a given size (e.g., the mean trade size). Each measure is computed for both the ask and bid side in the book and we include the imbalance between both sides at snapshot $t-1$ in our empirical specification as a potential predictor of short-term midquote returns over the interval between $t-1$ and t. We address predictability of these short-term returns on the basis of these book imbalances by considering the adjusted R^2 of the time series regression that is estimated for each stock and for each month in the sample. Using the R^2 as a metric to asses intraday return predictability in in line with previous research by, e.g., Chordia, Roll, and Subrahmanyam (2005) and Cao, Hansch, and Wang (2009).

We show that all our measures of book information can predict short-term returns and predictability is larger when using those measures that are limited to top-of-the-book information. As more information from deeper in the book is captured in a measure, the predictive ability of this measure decreases. This suggests that the most relevant information for future prices is concentrated at the top of the book, while deeper in the order book limit orders are more unrelated to future prices.

Having verified predictability we turn to the main contribution of our research, which is to address the sources of return predictability. Previous research discusses two potential explanations, but does not make an effort to disentangle between both. The first explanation is related to how traders adapt their order choice and starts from the observation that in a limit order market, traders can choose whether to submit a market order or a limit order, and furthermore can adjust the *aggressiveness* of their order. As shown in Parlour (1998), traders

[1] We also use ticks and the half-spread to define X but results are similar.

face a trade-off: while a limit order implies a better price, it also faces execution risk. Due to time priority rules, the probability of execution for a limit order decreases when there is already a large number of outstanding orders at its side of the book. By the same logic, when a large number of orders is available on the opposite side of the order, chances are higher that the trader will submit a limit order, since the probability that traders on the opposite side submit market orders increases (and thereby the probability of execution of his limit order). Parlour shows that even in the absence of asymmetric information and with a random arrival of trader types, systematic (and thus predictable) patterns in order flow emerge. Empirical evidence confirms such patterns: Ranaldo (2004) shows that a thicker (thinner) book on the same (opposite) side induces traders to submit more (less) aggressive orders. Furthermore, Pascual and Veredas (2009b) find that the frequency of patient traders increases with the thickness of the depth on the opposite side, while it decreases in the thickness of the depth on the same side. Patient traders base their order choice mostly on the state of the book at their side, while impatient traders pay more attention to the depth on the opposite side.[2] Not only newly arriving traders have to make an order choice decision. Many traders are present in the market for a longer time and employ dynamic order submission strategies, which means that they frequently cancel and resubmit orders. When they do so, they are likely to become increasingly aggressive insofar their previous orders remain unexecuted, and hence are more likely to switch to price-improving limit orders or market orders when the market is heavier on their side (or vice versa). In sum, the state of the book affects traders' order choice considerations, and thereby leads to short-term price changes. But any price effect that is purely due to order flow patterns that are predictable on the basis of order choice should be transitory (Goettler, Parlour, and Rajan, 2005). Indeed, if the fundamental value of the stock has not changed, arbitrageurs drive the price back to its fundamental value. Hence, if order choice and patterns are driving return predictability, we should observe no permanent effect of book imbalances. We refer to this first explanation as the 'order choice hypothesis'.

The second potential source for the return-imbalance relationship relies on informed traders submitting orders to the book, and hence we dub it the 'informed trading hypothesis'. Whether or not informed traders use limit orders or only rely on market orders as a part of their strategy is open to debate. For instance, Glosten (1994), Rock (1996) and Handa, Schwartz, and Tiwari (2003) argue that informed traders do not submit limit orders if their private information is short-lived, because there is a large probability that their limit orders would remain unfilled. Harris (1998) conjectures that the use of market orders

[2]Other theoretical and empirical work on the relation between the state of the order book and order choice includes Biais, Hillion, and Spatt (1995), Griffiths, Smith, Turnbull, and White (2000), Goettler, Parlour, and Rajan (2005), Cao, Hansch, and Wang (2008) and Duong, Kalev, and Krishnamurti (2009).

by informed traders depends on transaction costs on the one hand, and on the persistence of their informational advantage on the other hand. A number of theory models argue that informed traders can also submit limit orders. Kaniel and Liu (2006) find that, under certain conditions, informed traders are more likely to submit limit orders instead of market orders, even to an extent that limit orders convey more information than market orders. First, the higher the probability that the information is long-lived, the lower is the execution risk, favoring limit orders. Second, the larger the mispricing in the market, the higher are potential losses of non-execution, suggesting a greater favor towards market orders. Third, more uninformed traders increase the profitability of a limit order, but increase the profitability of a market order even more. Further, in the model of Roşu (2015), an informed trader decides between a limit and market order based on the magnitude of her information advantage. Traders with a large information advantage then opt for a market order, those with a smaller advantage submit a limit order. Also in the model of Goettler, Parlour, and Rajan (2009), informed traders tend to submit limit orders, but their liquidity supply is inverse related to (fundamental) volatility. Beber and Caglio (2005) find that informed agents act strategically and that their order submission strategies are partly dependent on market conditions. In circumstances where the probability of information based trading is high, they submit limit orders at a price further away in the book in order to hide their information. Through an experimental design, Bloomfield, O'Hara, and Saar (2005) find evidence that informed traders use both market and limit orders. They document a heavier use of market orders by informed traders at the beginning of the trading period, since mispricing is the largest then, and potential profits are greatest, consistent with predictions from Kaniel and Liu (2006). As prices are moving closer to their true market values, however, informed traders use more limit orders, because the additional expected profits from market orders decrease. Because the risk of being 'picked off' (adverse selection risk) is much smaller (or non-existent) for informed traders, this makes them natural liquidity suppliers. Hence, even the more distant levels may contain information.

If informed traders submit limit orders, then the state of the book resulting from these orders should reflect this information, and thus the book becomes more helpful in predicting future returns. Moreover, by incorporating their information into the limit order, prices become more and more efficient during the course of trading. Therefore, if asymmetric information is present, one should also observe a stronger and permanent price effect.

We address the question of what drives return predictability in three complementary ways. First, based on the reasoning above, we study predictability at short versus longer horizons. If returns are only predictable at short horizons, predictability is likely driven by short-term price pressure due to order choice considerations. If returns are also predictable over longer horizons based on the same limit order book information, this is akin to a permanent price effect and suggests that returns are predictable because the limit order book contains fundamental information. Second, we examine whether predictability varies over different periods in the trading day. The motivation for this approach is that informed traders are known to be more active at the beginning and near the end of the trading day (see, e.g., Admati and Pfleiderer, 1988; Garvey and Wu, 2009). Their activity at the beginning of the trading day is driven by the size of their informational advantage as overnight information increases asymmetric information. In addition, Bloomfield, O'Hara, and Saar (2005), Kaniel and Liu (2006) and Roşu (2015) predict that under these circumstances informed traders are more inclined to resort to market orders and less to limit orders. If informed traders thus cause returns to be predictable, predictability based on the book should be lower at the start and end of the trading day. Third, we perform panel data regressions where we regress the adjusted R^2 from the prediction models on a number of explanatory variables, including a measure of informed trading, quoted depth, volatility and trading volume.

Our results show that evidence is more in line with the order choice hypothesis. First, we find that for all our measures return predictability based on book information increases from 1 minute intervals up to 10 or 15 minute intervals. After that return predictability diminishes again for most measures, while it does not disappear. But there are remarkable differences between large and small stocks. For smaller stocks return predictability is quite low at short horizons, and then further increases up to 30 minutes. For large stocks we find the opposite pattern. These results suggest that there is both a transitory component to return predictability as well as a permanent component. The transitory component can be attributed to order choice considerations. Newly arriving traders are crowded out due to time and price priority rules: as more liquidity piles up on one side of the book traders on that side become more aggressive in order to obtain a satisfying execution probability. If they post a limit order at the end of the queue their risk of non-execution is relatively higher compared to jumping the queue and improving the best quoted price on their side of the book. To the extreme, traders observing a crowded book on their side may switch from limit orders to market orders in order to secure their trade. Similarly, traders on the opposite side become less aggressive as their execution probabilities are relatively more favorable. This kind of behavior can make returns predictable over a short horizon. Because the price changes are unrelated to the fundamental value of the security (informed) traders should jump in to revert prices back to their normal levels. The decrease in return predictability over longer horizons suggests that indeed some traders engage in strategies to remove transitory predictability. For more actively traded stocks return predictability decreases at a faster pace.

Second, we find no obvious pattern of return predictability throughout the trading day. When we divide the trading day into five subperiods of equal size, return predictability is larger in the second and third subperiod of the day, and the lowest for the fourth subperiod. For both the first and fifth subperiod predictability is 'intermediate', while the informed trading hypothesis predicts it should be the lowest.

Third, we show that predictability is cross-sectionally negatively related to informed trading, i.e. the presence of more informed traders in a stock decreases return predictability. Furthermore, return predictability is larger for stocks with deeper order books. This is a strong indication that return predictability based on book information is predominantly caused by temporary effects due to order choice considerations. On the one hand, informed trading is associated with a faster price discovery process. When prices temporary deviate from fundamental values they revert to normal levels faster when more informed traders are present. On the other hand, deeper order books make queue-jumping behavior relatively more attractive, as traders compete more heavily for order execution.

The remainder of the chapter is organized as follows. We start by explaining how information from the limit order book can be used to predict returns in Section 1.2. Next, we describe the dataset and institutional setting, and establish return predictability in our sample in Section 1.3. Subsequently, Section 1.4 investigates the drivers of the return predictability. Finally, Section 1.5 concludes.

1.2 Predicting Returns with Limit Order Book Information

1.2.1 Predicting Returns

To predict returns from order book information we use imbalances in liquidity measures between the bid and ask side. When the bid (ask) side is more liquid than the ask (bid) side this signals a heavier trading interest in the limit order book on the buy (sell) side, and therefore we expect positive (negative) returns.[3] We estimate the following time series model to evaluate predictability based on order book information using various imbalance measures:

$$r_t = \beta_0 + \beta_1 b_{t-1} + u_t. \tag{1.1}$$

with r_t the log return on the midquote m_t between time $t-1$ and t. We predict quoted prices from the order book instead of transaction prices because, according to our two explanations, these should be predictable from order book information. We use the midquote of the best bid and ask price because bid and ask prices are symmetrically affected. In addition, the midquote is often taken to be an estimate of the fundamental value of a security and therefore frequently used

[3]The use of imbalances in market variables that can capture the buying and selling behavior to predict short-term returns is not new (see, e.g., Chordia, Roll, and Subrahmanyam, 2002, 2005; Chung and Hrazdil, 2012). We use a similar methodology by regressing returns on lagged imbalances constructed from order book information.

as a benchmark, for instance in the estimation of trading costs. b_{t-1} is a measures of order book imbalance at the end of interval $t-1$, u_t is the error term. By using order book imbalances at $t-1$ we ensure that this information is available to market participants to include in their decisions. We evaluate predictability of returns by investigating the adjusted R^2 of our models as well as the estimated coefficients and their statistical significance. The higher the adjusted R^2 is, the more public information on the limit order book helps in predicting returns. Equation (1.1) is estimated per stock on a monthly basis. In other words, we use all snapshots t in a given month for a given stock. This results in a panel of adjusted R^2 for the stocks and months in our sample. This allows us to explore both the cross-sectional and time series variation in predictability later in Section 1.4.

Furthermore, following Chordia, Roll, and Subrahmanyam (2008) and Chung and Hrazdil (2012) we use raw returns r_t as our dependent variable in our main analysis. In contrast, Cao, Hansch, and Wang (2009) use innovations in returns from an autoregressive model to capture the unpredictable part in the returns, while Chordia, Roll, and Subrahmanyam (2005) include lagged returns in their model. Our results remain qualitatively unchanged when both these approaches are used and are available upon request. We favor the use of the simplest version of the model because it is most easy to interpret. We show that market participants can predict future price movements by using a single metric that captures order book information.

We need to make a choice about the frequency at which these regressions are estimated, i.e. a choice about the length of the interval $[t-1, t]$. The choice of the interval length matters since predictability is likely to differ for different interval lengths. Chordia, Roll, and Subrahmanyam (2005) find that market order imbalances can predict returns over smaller time intervals of less than 30 minutes, but as the interval length increases predictability diminishes. A comparison of model estimates using different interval lengths also sheds more light on the question of whether imbalances in the limit order book are caused by informed traders submitting limit orders. Therefore we also investigate different sampling frequencies: 1 minute, 2 minutes, 5 minutes, 10 minutes, 15 minutes and 30 minutes, respectively dividing the trading day into 510, 255, 102, 51, 34 and 17 intervals (since trading in our sample takes place between 9 a.m. and 5.30 p.m.).

1.2.2 Limit Order Book Information

For our analysis, we require a measure that captures the state of the limit order book at a given point in time. For this, we first take snapshots of the limit order book every x minutes, a particular snapshot is represented by subscript t. Below, we consider snapshots every x = 1, 2, 5, 10, 15 and 30 minutes. Next, we define three measures of the state of the limit order book at t. Important to stress is that our measures not only capture the state at the best prices of the limit order book, but also deeper into the book. We then focus in particular on the imbalance between the ask and bid side of the limit order book. Given that we want to study how the public limit order book, as observed by traders, helps to predict returns, we do not include hidden depth in the imbalance measures.

Our first measure is the slope of the limit order book, as defined in Naes and Skjeltorp (2006). Let $p^{ask}_{j,t}$ denote the ask price (in euro) at level $j = 1, ..., J$ in the limit order book at snapshot t, and $q^{ask}_{j,t}$ the number of shares available at that ask price. m_t denotes the midquote. Symmetrically, $p^{bid}_{j,t}$ and $q^{bid}_{j,t}$ denote the bid prices and available shares. The slopes at t, taking into account the first L levels in the book at the ask side $\left(Slope^{ask}(L)_t\right)$ and bid side $\left(Slope^{bid}(L)_t\right)$ are then computed as

$$Slope^{ask}(L)_t = \frac{1}{L}\left\{\frac{\ln\left(q^{ask}_{1,t}\right)}{\frac{p^{ask}_{1,t}}{m_t}-1} + \sum_{l=1}^{L}\frac{\frac{\ln\left(q^{ask}_{l+1,t}\right)}{\ln\left(q^{ask}_{l,t}\right)}-1}{\frac{p^{ask}_{l+1,t}}{p^{ask}_{l,t}}-1}\right\} \quad (1.2)$$

$$Slope^{bid}(L)_t = \frac{1}{L}\left\{\frac{\ln\left(q^{bid}_{1,t}\right)}{\frac{p1,t^{bid}}{m_t}-1} + \sum_{l=1}^{L}\frac{\frac{\ln\left(q^{bid}_{l+1,t}\right)}{\ln\left(q^{bid}_{l,t}\right)}-1}{\frac{p^{bid}_{l+1,t}}{p^{bid}_{l,t}}-1}\right\} \quad (1.3)$$

The order book slope can be interpreted as an elasticity, i.e. the slope measures how order book depth (shares that are supplied or demanded) changes as the price changes. Larger slopes indicate more liquid order books. The scaled imbalance between the bid and ask side of the limit order book slopes $ImbSlope(L)_t$ is defined as

$$ImbSlope(L)_t = \frac{Slope^{bid}(L)_t - Slope^{ask}(L)_t}{Slope^{bid}(L)_t + Slope^{ask}(L)_t} \quad (1.4)$$

A positive slope imbalance indicates a more liquid bid side and hence more patient buying interest than selling interest. Note that by varying L, the number of price levels in the limit order book taken into account to measure the slope, *ImbSlope* allows to address the imbalance both close to the best prices, and further in the limit order book. In our analysis we include $L = 1, 3$ and 5.

Our second measure is the *Depth* (X) measure from Degryse, de Jong, and van Kervel (2015). It captures depth available in the limit order book in an interval around the midpoint. More specifically, using the same notation as above, the *Depth* (X) measure for the ask and bid side of the limit order book is

$$Depth^{ask}(X)_t = \sum_{j=1}^{J} p^{ask}_{j,t} q^{ask}_{j,t} \mathbb{1}\left(p^{ask}_{j,t} < m_t(1+X)\right) \quad (1.5)$$

$$Depth^{bid}(X)_t = \sum_{j=1}^{J} p^{bid}_{j,t} q^{bid}_{j,t} \mathbb{1}\left(p^{bid}_{j,t} > m_t(1-X)\right) \quad (1.6)$$

where $\mathbb{1}(.)$ is an indicator function that is one if the expression in the brackets is satisfied, and zero otherwise. Our scaled imbalance measure $ImbDepth\,(X)_t$ at snapshot t is then defined as

$$ImbDepth\,(X)_t = \frac{Depth^{bid}\,(X)_t - Depth^{ask}\,(X)_t}{Depth^{bid}\,(X)_t + Depth^{ask}\,(X)_t} \tag{1.7}$$

To compute $Depth(X)$, we need to define the interval of size X around the midquote that we want to take into account. We consider intervals around the midpoint from 30, 70 and 100 basis points (b.p.).[4] The intuition for this measure is that a low value for X captures depth at or close to the best quotes, while larger values take into account a larger part of the limit order book. When only depth up to a certain level of the order book can be observed in real-time this constitutes a maximum boundary for each depth measure that we can calculate. In our sample this is the fifth level (see Section 1.3). For highly liquid stocks that have a tight bid-ask spread and for which the distance between consecutive prices is relatively short the maximum bound is encountered relatively more often, and for lower values of X. But for illiquid stocks the maximum bound may never be encountered. By contrast, $Depth\,(X)$ is often zero for low levels of X, which leaves $ImbDepth\,(X)_t$ to be either undefined or zero. We choose to set $ImbDepth\,(X)_t$ to zero for those observations, since the (zero) depth is in fact balanced and it allows us to retain the complete time series. As the number of zero (or undefined) depth imbalances for a stock-month combination grows too large, however, results become less meaningful. We therefore exclude estimation results of a given stock-month combination from the sample for which the fraction of zero depth imbalances exceeds a threshold (set to 50%, but results are similar when using either no threshold, a 25% threshold or a 75% threshold).

Our third measure is based on the cost of a round trip trade CRT in Irvine, Benston, and Kandel (2000). We compute the cost (in euro) of a buy transaction of a particular size at snapshot t, $CT^{ask}\,(Size)$, as such capturing the ask side of limit order book. The same is done for the cost a sell transaction of a particular size $CT^{bid}\,(Size)$. We then define the scaled imbalance measure $ImbCT\,(Size)_t$ at time t as follows:

$$ImbCT\,(Size)_t = \frac{CT^{ask}\,(Size)_t - CT^{bid}\,(Size)_t}{CT^{ask}\,(Size)_t + CT^{bid}\,(Size)_t} \tag{1.8}$$

We subtract the cost of a trade on the bid side from the cost of a trade on the ask side instead of the other way around, as with our previous measures. The reason is that CT is a measure of illiquidity, while both the $Slope$ and $Depth\,(X)$ are measures of liquidity. When we instead add CT^{bid} and CT^{ask} we get the cost

[4] We also considered two other ways to define the interval size X, measured in ticks and half-spreads. Results are qualitatively similar using these alternative definitions and are available upon request.

of a round trip trade CRT, a frequently used measure of order book liquidity. We consider trade sizes defined in different ways, more specifically the mean trade size of the stock in the month in which the snapshot was taken, and the mean trade size + 1 and 2 standard deviations. By including small and large trade sizes, we again take into account the limit order book close to the best quotes, as well as further way from the best prices. When the trade size exceeds depth available on the visible quotes we assume, for the sake of simplicity, the remainder of the trade can be executed at the final visible quote.

1.3 Data

1.3.1 Institutional Setting

This study uses intraday data of stocks that are listed on the Spanish stock exchange, which is a purely order-driven market. Its trading activity is managed through the electronic trading platform Spanish Stock Market Interconnection System (SIBE). Investors submit their orders through brokers who are provided with real-time information on trading activity and the state of the limit-order book by SIBE's Dissemination Information System (DIS). Continuous trading takes place from 09.00 a.m. until 05.30 p.m. and call auctions determine the opening and closing price. The minimum order size is 1 share and the tick size is dependent on the trade price. It is €0.05 for stocks with a price above €50 and €0.01 for stocks with a lower price (all stocks in our sample). Three types of updates in the limit order book can be distinguished: new orders, order modifications and order cancellations. Market orders are executed against the best prices on the opposite side of the book and walk up or down the book until they are completely filled. Market-to-limit orders are like market orders, but do not walk up or down the book when the depth at the first best price is completely used. Instead, they are stored at that price as a regular limit order. Limit orders are recorded in the book at their limit price and can only be executed at that price or better. Priority of execution is based on order submission time. The unmatched limit orders summarize into the state of the book. The dataset contains the book from the first until the fifth level, except for the invisible part of the depth from iceberg orders.

1.3.2 Sample Description

Our sample contains all stocks that were part of the Open Market (Mercado Open) and the New Market (Nuevo Mercado) from January until December 2003. Stocks from the fixing markets are not considered because they do not trade in a continuous limit order market. Stocks in the Latibex-segment, which contains Latin-American stocks that are cross-listed in Spain, are discarded as well. We remove stocks that have an average daily number of trades that is less than 10, and an average price below €0.5. Our final sample consists of 88 stocks. The sample period covers all trading days for the stock during 2003, which is 250 days for most stocks.

We combine two data files in our analysis. One file contains data on the limit order book and shows all updates of the first five levels in the book for each stock in the sample, time-stamped to the nearest hundredth of a second. Every update contains the five best prices at both sides of the book and their respective depths. A second file contains all trades executed during the continuous trading session. The trading data show price and size of each trade and are time-stamped to the nearest second. Pre-opening or post-closing orders are not included since the trading mechanism during this period is different from the one during the trading day. Both book updates and trades are indexed. The index numbers and time stamps allow for a perfect matching of trade and order book data. We only consider observations during the continuous trading session.

Table 1.1 Panel A displays some descriptive statistics for the sample. The cross-sectional mean, standard deviation, median, quantiles and 5th and 95th percentile for a selection of daily stock characteristics (pooled across stocks and days) is shown.

Our sample contains small as well as large firms, with a 5th and 95th percentile of daily market capitalization at 58 million euro and 17 billion euro respectively. The median firm has a market value of 1.1 billion euro and a stock price of 10 euro. There is also considerable variation in trading activity. The 5th percentile of the number of transactions (euro volume traded) is merely 7 trades (17,100 euro traded). By contrast, the 95th percentile is at 1,672 transactions (64,947,610 euro traded) per day. The median stock has 82 transactions per day with 828,880 euro transacted. The low levels of trading and quoting activity (an average of 2,037 and median of 308 book updates per day) indicate that these data are before the advent of high-frequency trading. For about 5% of the observations the daily spread is close to its minimum value of 1 tick. The mean (median) spread is 5.79 (3.24) ticks, or 0.60% (0.40%) when measured relative to the midquote. The PIN is a measure of informed trading discussed in Subsection 1.4.3.

Table 1.1 Panel B displays descriptive statistics for the liquidity measures from our sample, based on daily observations of these measures (with the daily measures being time-weighted averages across the trading day). The $Slope$ of the order book is the average of the bid and ask slope, and is decreasing as more information from deeper in the order book is incorporated. The $Depth$ (X) measure is defined at levels of depth relatively close to the midquote, as well as levels of depth further down the order book. By definition, $Depth$ (X) increases in X. $Depth$ (X) has observations of zero depth for X close to the midquote for stocks with relatively large spreads. The mean (median) $Depth$ (X) at 30 basis points is 253,150 euro (30,820 euro), while the mean (median) at 100 basis points is

Table 1.1: Descriptive Statistics

This table presents descriptive statistics for each of the 88 stocks in our sample: the mean, standard deviation, median, 5th percentile, first quantile, median, third quantile and 95th percentile of selected variables based on daily observations. Panel A presents statistics on the market value of the stock and order book and trade characteristics. Volume and trade size are measured in thousands of euro. The number of updates is the total number of order book updates during a day: trades, limit order submissions, cancellations and modifications. Market value is measured in millions of euro. The percentage spread is measured relative to the midquote. PIN is a measure of the probability of informed trading, and its (monthly) estimation is detailed in . Panel B shows statistics on order book state (liquidity) measures. *Slope* is an elasticity-based measure of order book liquidity that combines information from different levels in the order book. It is the average of the slopes of the bid and ask side. Its value is divided by 100. *Depth* (X) combines bid and ask depth expressed in thousands of euro and measured as X basis points away from the midquote. CRT is the cost of a round trip transaction of a given size, either (1) the mean trade size, (2) the mean trade size plus one standard deviation, (3) the mean trade size plus two standard deviations. It is expressed relative to the total euro size of the transaction. *Spread*, *Slope* , *Depth* (X) and CRT are time-weighted over the trading day. Price is volume-weighted over the trading day. Panel C (Panel D)presents similar statistics as Panel A (Panel B), but then for 3 subsamples based on market capitalization tertiles.

Panel A: Stock Characteristics

		Mean	St. Dev.	P5	Q1	Median	Q3	P95
Market Value		3,762.21	8,050.63	58.48	216.27	1,134.53	2,955.07	17,457.12
Price		12.44	9.84	1.27	4.83	10.14	17.14	34.66
Volatility		1.25%	0.81%	0.41%	0.73%	1.04%	1.52%	2.72%
Volume		12,357.96	42,190.72	17.10	131.87	828.88	4,986.17	64,947.61
Nr Trades		346.37	729.89	7.00	22.00	82.00	338.00	1,672.00
Trade Size		16.19	31.93	1.73	4.89	9.51	19.36	50.43
Nr Updates		1,174.87	2,039.18	21.00	75.00	308.00	1,382.00	5,452.80
Spread	tick	5.79	8.92	1.04	1.77	3.24	6.67	18.35
	%	0.60%	1.09%	0.11%	0.22%	0.40%	0.77%	1.83%
PIN		22.21%	11.18%	9.84%	15.05%	19.90%	25.82%	42.95%

Panel B: Order Book State/Liquidity Measures

		Mean	St. Dev.	P5	Q1	Median	Q3	P95
	l.1	64.39	54.24	8.86	23.03	47.66	89.00	181.94
Slope	l.3	22.30	18.37	3.40	8.22	16.65	30.78	61.87
	l.5	13.55	11.07	2.11	5.07	10.16	18.69	37.33
	30 b.p.	253.15	921.70	0.00	3.26	30.82	154.18	1,043.62
Depth (X)	70 b.p.	464.04	1,570.31	0.00	27.58	105.61	308.58	1,568.07
	100 b.p.	504.14	1,610.78	3.41	46.11	140.66	346.25	1,757.91
	m	0.87%	1.10%	0.14%	0.29%	0.57%	1.09%	2.53%
CRT	m+sd	1.40%	1.66%	0.25%	0.51%	0.96%	1.76%	3.99%
	m+2sd	1.71%	1.95%	0.32%	0.66%	1.20%	2.13%	4.66%

Table 1.1 continued.

Panel C: Stock Characteristics by Subsample

		Mean	P5	Median	P95	Mean	P5	Median	P95	Mean	P5	Median	P95
		Small Cap				Mid Cap				Large Cap			
Market Value		259.97	21.39	111.15	810.36	1,270.06	307.90	1,096.72	4,118.19	9,840.90	1,421.34	4,277.69	36,764.37
Price		6.97	0.53	4.17	17.25	12.81	3.06	11.11	25.43	17.53	4.48	14.97	40.71
Volatility		1.35%	0.44%	1.13%	3.02%	1.12%	0.31%	0.92%	2.58%	1.29%	0.47%	1.06%	2.70%
Volume		299.17	6.64	94.87	1,218.81	2,007.36	47.57	757.06	7,039.29	35,120.91	684.76	8,115.76	179,606.98
Nr Trades		37.63	4.00	20.00	119.95	170.35	9.00	77.00	503.00	837.01	70.55	441.00	3,252.50
Trade Size		7.69	1.01	4.35	21.77	13.72	2.75	9.32	31.07	27.22	5.95	21.36	64.44
Nr Updates		131.51	15.00	60.00	455.00	509.83	30.00	271.00	1,518.50	2,905.60	315.00	1,943.00	9,044.05
Spread	tick	6.77	1.00	3.50	24.25	6.71	1.14	3.61	21.73	3.85	1.05	2.59	10.22
	%	1.08%	0.32%	0.86%	2.58%	0.51%	0.17%	0.39%	1.19%	0.23%	0.09%	0.19%	0.55%
PIN		25.88%	10.05%	23.02%	57.12%	21.57%	8.97%	19.87%	40.02%	18.71%	10.33%	17.11%	32.36%
AS		33.43%	1.70%	27.04%	82.05%	27.33%	2.56%	21.07%	76.29%	33.28%	1.34%	26.87%	82.88%

Panel C: Order Book State/Liquidity Measures by Subsample

		Mean	P5	Median	P95	Mean	P5	Median	P95	Mean	P5	Median	P95
		Small Cap				Mid Cap				Large Cap			
	l.1	24.33	6.24	19.87	56.74	53.39	13.67	48.25	112.91	115.80	38.69	106.04	217.18
Slope	l.3	8.56	2.40	7.06	19.73	18.67	5.52	16.86	38.80	39.78	13.65	36.72	73.66
	l.5	5.23	1.51	4.33	11.95	11.36	3.47	10.25	23.54	24.04	8.31	22.25	44.36
		Small Cap				Mid Cap				Large Cap			
	30 b.p.	41.23	0.00	2.28	44.45	57.91	0.00	31.74	202.62	684.48	26.44	243.11	2,865.67
Depth (X)	70 b.p.	87.50	0.00	18.92	140.98	152.34	18.06	105.49	430.84	1,194.94	98.15	410.04	5,722.36
	100 b.p.	108.27	0.42	34.43	191.04	187.10	32.82	139.94	490.56	1,262.66	129.04	442.02	5,797.44
		Small Cap				Mid Cap				Large Cap			
	m	1.58%	0.51%	1.25%	3.52%	0.74%	0.23%	0.57%	1.75%	0.30%	0.12%	0.25%	0.68%
CRT	m+sd	2.40%	0.78%	1.94%	5.31%	1.28%	0.40%	0.99%	3.08%	0.53%	0.17%	0.44%	1.17%
	m+2sd	2.88%	0.96%	2.35%	6.21%	1.58%	0.52%	1.26%	3.56%	0.67%	0.22%	0.58%	1.44%

504,140 euro (140,660 euro). The cost of a round-trip transaction is the sum of the cost to buy at the ask side CT^{ask} and the cost to immediately sell at the bid side CT^{bid}, for a transaction of a given size and it is expressed relative to the size. The mean (median) cost of a round-trip transaction of mean size is 0.87% (0.57%) of the value of the transaction. This is a little bit more than the relative spread, suggesting that the mean trade size consumes on average more depth than available at the best quotes. It increases up to 1.71% (1.20%) for a transaction size equal to the mean plus 2 standard deviations.

We examine differences between subsamples of our stocks more closely in Panels C and D of Table 1.1. We distinguish between three subsamples on the basis of the free-float market capitalization: small cap, mid cap and large cap.[5] The mean (median) market value in the small cap sample is 260 million euro (111 million euro), while it is 1.27 billion euro (1.097 billion euro) for mid cap stocks and 9.841 billion euro (4.278 billion euro) for the large cap subsample. There is also quite a large difference in trading activity and liquidity. The average daily volume is 299,170 euro for small cap stocks, with only about 38 trades a day and 132 limit order book updates. For large cap stocks the average daily trading volume is much higher at 35,120,910 euro. The number of trades is 837 per day on average, and the limit order book is updated 2,906 per day – order submissions, cancellations and modifications. The average relative spread is 1.08% of the midquote for small cap stocks and 0.23% for large cap stocks. Depth at 30 basis points (100 basis points) around the midquote is 41,230 euro (108,270 euro) for small cap stocks and 684,480 euro (1,262,660 euro) for large cap stocks. Values for market capitalization, trading activity and liquidity for mid cap stocks are in between that of small cap stocks and large cap stocks. However, they appear to be closer to that of small cap stocks than mid cap stocks. In fact, the market capitalization and trading activity of our sample is dominated by a handful of stocks, while the majority of stocks is rather illiquid. We exploit the differences between different stocks in Section 1.4.

Summary statistics of the imbalances in the state of the book are presented in Panel A of Table 1.2, based on daily time-weighted measures. For all measures the ask side appears more liquid than the bid side, as both the mean and median are slightly negative. The mean (median) imbalance in the $Slope$ of the order book using the first level is -0.34% (-0.21%). The order book is on average (median) 1.51% (0.00%) deeper on the ask side compared to the bid side at 30 basis points around the midquote, while the cost of a trade is on average (median) 0.98% (0.22%) higher on on the bid side than on the ask side. Furthermore, the more information from deeper in the book is used to calculate imbalance measures, the greater becomes the imbalance between the ask and bid side. The average (median) imbalance in $Depth$ (X) amounts to -2.94% (-2.19%) at 100

[5]For each month we sort stocks into subsamples of equal size based on their free-float market capitalization.

Table 1.2: Descriptive Statistics: Imbalances in the Limit Order Book

Panel A presents statistics on the the imbalance in the state (liquidity) of the order book between the bid and ask side (in percentage points), using the following three measures of liquidity. *Slope* is an elasticity-based measure of order book liquidity that combines information from different levels in the order book. *Depth* (X) combines bid and ask depth expressed in thousands of euro and measured as X basis points away from the midquote. *CT* is the cost of a transaction of a given size, either (1) the mean trade size, (2) the mean trade size plus one standard deviation, (3) the mean trade size plus two standard deviations. The imbalance in order book *Slope* is defined in Equation (1.4), the *Depth* (X) imbalance is defined in Equation (1.7) and the *CT* imbalance is defined in Equation (1.8). Imbalances are time-weighted throughout the trading day. Panel B presents correlations between daily order book state imbalance measures.

Panel A: Imbalances in the Order Book State									
		Mean	St. Dev.	P5	Q1	Median	Q3	P95
Slope	1.1	-0.34	9.41	-10.97	-2.99	-0.21	2.41	9.49
	1.3	-0.56	5.76	-9.79	-2.67	-0.25	1.91	7.46
	1.5	-0.60	5.56	-9.66	-2.63	-0.28	1.81	7.11
Depth (X)	30 b.p.	-1.51	18.65	-32.93	-9.77	0.00	6.65	27.82
	70 b.p.	-2.33	26.11	-47.84	-16.79	-0.88	12.73	40.61
	100 b.p.	-2.94	27.70	-50.84	-19.00	-2.19	13.35	42.82
CT	m	-0.98	12.05	-21.63	-5.03	-0.22	3.71	16.92
	m+sd	-1.27	16.76	-30.64	-9.36	-0.55	7.28	25.46
	m+2sd	-1.26	17.81	-32.20	-10.63	-0.67	8.42	27.23

Panel B: Correlations in Imbalances in the Order Book State										
		Slope			Depth (X)			CT		
		1.1	1.3	1.5	30 b.p.	70 b.p.	100 b.p.	m	m+sd	m+2sd
Slope	1.1	1.00								
	1.3	0.44	1.00							
	1.5	0.43	0.97	1.00						
Depth (X)	30 b.p.	0.38	0.58	0.57	1.00					
	70 b.p.	0.36	0.60	0.59	0.65	1.00				
	100 b.p.	0.33	0.59	0.59	0.53	0.87	1.00			
CT	m	0.48	0.72	0.70	0.59	0.60	0.57	1.00		
	m+sd	0.33	0.62	0.63	0.59	0.70	0.70	0.69	1.00	
	m+2sd	0.26	0.54	0.57	0.53	0.69	0.71	0.55	0.93	1.00

basis points around the midquote. Imbalances in the *Slope* of the book take less extreme values, with a range from -9.66% to 7.11% between the 5th and 95th percentile when using the 5 best levels, compared to a -50.84% to 42.82% for *Depth* (100). Do note that these are daily time-weighted imbalances and at higher frequencies the imbalances can take on more extreme values.

Correlations between the imbalances are shown in Panel B of Table 1.2. All imbalance measures are positively correlated. Correlations between intermediate and deeper order book imbalance measures are in general quite large: 0.97 for the *Slope*, 0.87 for *Depth* (X) and 0.93 for CT. Correlation between the top-of-the-book imbalances and intermediate imbalances are lower, with 0.44 for the *Slope*, 0.65 for *Depth* (X) and 0.69 for CT.

1.3.3 Book Imbalances and Return Predictability

Table 1.3 shows that returns are statistically predictable using limit order book information over 5 minute intervals. Recall that all models are estimated on a per stock per month basis. We exclude, for each type of measure, all stock-month combinations for which imbalances are not defined or equal to zero for more than 50% of the observations. This leaves us with a full sample for the $Slope$ measure, 714 stock-month combinations for the $Depth\ (X)$ measure and 863 stock-month combinations for the CT measure. The first column present the average coefficient of the imbalance measure, β_1 in Equation (1.1) over all regressions, while the second column presents the average t-statistic, which ranges between 4.57 and 7.79, depending on the model, with standard errors adjusted using the Newey-West correction for serial correlation. The third column shows the percentage of regressions for which the imbalance measure was significant at the 5% level. Depending on the measure used, limit order book imbalances are significantly related to future returns for 88% to 93% of the relevant sample. As our main determinant of predictability we investigate the adjusted R^2 in columns 4 to 9. The mean (column 4) varies between 1.46% and 2.77%, the median (column 7) between 1.30% and 2.36%. Furthermore, columns 5 and 9 present the 5th and 95th percentile. For the 5th percentile estimates are in the range -0.03% - 0.17%, wile for the 95th percentile the range is 3.40% - 8.25%. The values are in line with previous research on intraday return predictability using past order flow or order book information (see, e.g., Chordia, Roll, and Subrahmanyam, 2005, 2008; Cao, Hansch, and Wang, 2009; Chung and Hrazdil, 2012). Note that there is considerable variation in adjusted R^2 across the sample.

Furthermore, measures that use information from deeper in the book have lower predictive power than those that use information closer to the best prices. For instance, the mean adjusted R^2 from the $Depth\ (X)$ measure declines from 2.57% at 30 basis points to 1.46% at 100 basis points. As more information from deeper in the limit order book is aggregated in these imbalance measures the adjusted R^2s as well as the t-statistics and thus the number of statistically significant imbalances decreases. Orders that are submitted deeper in the order book appear to have a weaker relation to future price movements. When aggregated together with more informative top-of-the-book orders the overall predictive ability of the limit order book information decreases. The stronger relationship between future returns and top-of-the-order-book imbalances also holds for the $Slope$ and CT, although less outspoken. For the CT the decline using deeper order book information is not for all stock-months in the sample.

Our result that top-of-the-book information is most relevant confirms previous findings from Cao, Hansch, and Wang (2009). Orders further down the order book contain few relevant information with regard to the direction and size of future short-term price movements, above what is included in orders at the best levels of the book. There are two potential explanations for this finding, related

Table 1.3: Return predictability at 5 minute intervals

This table presents estimates of the different models from Equation (1.1). The independent variable is the return measured during a 5 minute interval (expressed in basis points). Explanatory variables are imbalances in the order book as defined in Subsection 1.2.2. The models are estimated on a per stock and per month basis for the 88 stocks in our sample, for a total of 1,056 stock-months. When an imbalance measure is not defined or only has values that are different from zero for less than 50% of the observations, we do not include the stock-month in our sample for that type of measure. For the slope measure we have results for the full stock-month sample, for the Depth (X) measure we have 714 stock-months (67.61% of the sample), for the CT measure we have 863 stock-months (81.72% of the sample). The first column shows the average coefficient, the second column shows the average t-statistic based on Newey-West HAC standard errors. The third column shows the percentage of regressions for which the coefficient of the imbalance is significantly positive at the 5% level. Columns 4 to 9 show distributional characteristics of the adjusted R^2 that are estimated: mean, quartiles, 5th and 95th percentiles.

		Coeff	t-stat	%sign	Adj R^2					
					Mean	P5	Q3	Median	Q3	P95
Slope	l.1	20.00	7.79	89.87%	2.77%	0.04%	0.74%	1.58%	2.79%	8.25%
	l.3	22.99	5.35	87.69%	2.39%	0.13%	0.82%	1.83%	3.30%	6.33%
	l.5	22.32	5.31	87.69%	2.32%	0.08%	0.77%	1.77%	3.22%	6.16%
Depth (X)	30 b.p.	5.86	6.39	92.86%	2.57%	0.17%	1.13%	2.36%	3.68%	5.64%
	70 b.p.	6.21	5.16	91.74%	1.90%	0.13%	0.92%	1.71%	2.63%	4.25%
	100 b.p.	6.27	4.57	89.92%	1.46%	0.10%	0.69%	1.30%	1.98%	3.40%
CT	m	8.68	5.37	91.66%	2.26%	-0.01%	1.01%	2.12%	3.35%	5.05%
	m+sd	8.12	5.64	90.27%	2.44%	-0.02%	0.91%	2.07%	3.74%	5.98%
	m+2sd	8.13	5.28	88.76%	2.14%	-0.03%	0.73%	1.81%	3.23%	5.33%

to the two potential sources we put forward to explain why order imbalances can predict returns. The first is related to informed trading in the limit order book. If predictability is caused by informed limit order submissions *and* the top of the order book is shown to be a better predictor of future returns compared to deeper levels of the order book, then informed traders submit more orders to the top of the book. Unfortunately we have little guidance on where in the limit order book informed traders are more likely to submit orders. Informed traders may prefer the top of the order book to increase their execution probability. But they may also prefer to hide their information deeper in the book. A second explanation is related to the order choice hypothesis. When patient traders decide on their order aggressiveness based on the state of the limit order book, their decision has the largest impact on prices when they switch from a limit order at the first level (the top of the order book) to a more aggressive order (or vice versa). Therefore imbalances at the top of the order book should be stronger related to future price movements than imbalances deeper in the order book, at least in the short run. We further investigate the potential sources of predictability using order book information in the following section.

1.4 What Explains Return Predictability on the Basis of the Limit Order Book?

We now address the question of what drives return predictability. Our main question of interest is whether return predictability is due to informed trading in the limit order book and thus whether imbalances in the limit order book carry fundamental information (the informed trading hypothesis), or whether returns are predictable because imbalances in the limit order book cause traders to generate short-term price pressure (the order choice hypothesis). We address this question by conducting three analyses. In the first subsection, we compare predictability over different time horizons. In the second subsection intraday differences in predictability are considered. Finally, the last subsection looks at determinants of predictability in a cross-sectional framework.

1.4.1 Predictability at Short versus Longer Horizons

In a first analysis, we study whether there is a difference in return predictability over different horizons. Chordia, Roll, and Subrahmanyam (2005) find that lagged market order imbalances lose their predictive ability over longer horizons, as the coefficient estimate, t-statistic and adjusted R^2 converge to zero. They conjecture that autocorrelated market order imbalances cause the specialist to alter his prices because of inventory concerns. Arbitrageurs are able to estimate order imbalances and their influence, but it takes a few minutes. As a reaction, they engage in countervailing trades, which removes predictability of returns. Chordia, Roll, and Subrahmanyam (2005) argue that these countervailing trades make the market efficient after a period of 30 minutes or less. An implication of their finding that lagged order imbalances cannot predict returns at longer horizons, is that informed traders do not engage in predictable order splitting, at least with regard to market orders. Market order imbalances do not carry any fundamental information on future prices, i.e. any information is instantaneously incorporated.

In order to evaluate whether order book state imbalances are informative, we estimate our models from Equation (1.1) for different time horizons and compare the results. We take snapshots of the book and calculate returns over intervals of 1, 2, 5, 10, 15 and 30 minutes. Results are presented in Table 1.4. Panel A shows a summary of the results based on the full stock-month sample, using imbalances in the order book slope, depth and the cost of a trade as a predictor (for each type of measure we show results of 3 versions to compare the top of the order book with the deeper book). As before, the average of the estimated coefficients and t-statistics together with the accompanying number of regressions with significant coefficients at the 5% level are shown in the first four columns. The last two columns show the mean and median adjusted R^2s. The adjusted R^2, our main measure of predictability, first increases with the time horizon, but subsequently drops to lower levels. The initial increase suggest that not all price relevant information contained in the order book is instantaneously reflected into prices, but the drop in adjusted R^2s also shows that at least a part of the effect is transitory in nature. The transitory price effect of order book imbalances is because traders choose their order aggressiveness on the basis of the current state of the order book. Because traders' aggressiveness is correlated with price and quote changes order book imbalances can predict future returns. When price movements are unrelated to the fundamental value of the security arbitrageurs will step in and submit orders against the direction of transitory price movements, thereby decreasing predictability of returns over longer horizons. As shown by Chordia, Roll, and Subrahmanyam (2005) it may take time for arbitrageurs to act on temporary price changes and remove predictability.

When comparing top-of-the-book measures with deeper order book measures notice that predictability, as measured by the adjusted R^2, is in general higher for the former measures for most intervals. The top of the book seems to contain most useful information on future price changes. The exception is the CT measure for which predictability is in general higher for the intermediate levels of order book information. When looking at the pattern we observe a difference in the decrease of predictability over time. In particular for the $Depth\ (X)$ measure the drop in predictability is less sharp when more information from deeper in the limit order book is used. The mean and median adjusted R^2 decreases between 5 and 10 minutes for $Depth$ (30), but only decrease between 15 and 30 minutes for $Depth$ (100) and remains at relatively high levels. There are two potential explanations. First, when informed traders submit more orders deeper in the order book, limit order book measures containing more information from deeper in the order book should contain more information, and hence bear a less transitory relation to future returns. Second, the state of the order book at the top is likely to be most relevant to impact traders' order submission decisions at the top of the book, and therefore also the best bid and ask prices. The deeper state of the order book should only affect order submission behavior that is *not* at the top. However, by incentivizing traders deeper in the order book to be more or less aggressive, indirectly also the state of the order book at the top is affected after a while, and finally order submission behavior at the top, and hence prices – but with a lag.

Note that while the magnitude of the coefficients is increasing over time horizons, the statistical significance decreases. For instance, for the $Slope$ the average t-statistic is monotonically decreasing over all intervals from a range of 8.17 - 8.59 for 1 minute intervals to 2.44 - 2.80 for 30 minute intervals. There is also a sharp drop in the number of regressions for which the order book slope is significant from 15 minute to 30 minute intervals, from 80.97% - 82.48% to 59.85% - 62.97%. This decrease in statistical significance is at least partially determined by the reduced power of the test. As the number of observations decreases (by a factor 30) so does the estimation error increase.

We now turn to differences in subsamples based on market capitalization tertiles. Panel B of Table 1.4 shows that there are remarkable differences in mean and median adjusted R^2 between small cap stocks and large cap stocks. For small cap stocks predictability is the lowest at high frequencies and then in-

Table 1.4: Return predictability at different horizons

This table presents estimates of the different models from Equation (1.1) for different time horizons [t-1,t]. The independent variable is the return measured during time intervals of 1, 2, 5, 10, 15 or 30 minutes (expressed in basis points). Explanatory variables are imbalances in the order book as defined in Subsection 1.2.2. The models are estimated on a per stock and per month basis for the 88 stocks in our sample, for a total of 1,056 stock-months. When an imbalance measure is not defined or only has values that are different from zero for less than 50% of the observations, we do not include the stock-month in our sample for that type of measure. For the slope measure we have results for the full stock-month sample, for the Depth (X) measure we have 715 stock-months (67.71% of the sample), for the CT measure we have 863 stock-months (81.72% of the sample). In Panel A we present a summary of the results based on the total sample in five columns for different imbalance measures. The first column shows the average coefficient, the second column shows the average t-statistic based on Newey-West HAC standard errors. The third column shows the percentage of regressions for which the coefficient of the imbalance is significantly positive at the 5% level. The fourth column shows the mean adjusted R^2 and the median is shown in the fifth column.
In Panel B we present the mean and median adjusted R^2 for the same models as in Panel A, but we break down results according to the market capitalization of the stocks. Small cap stocks are those in the first tertile, mid cap stocks are those in the second tertile, and large cap stocks are those in the third tertile according to the size of their free float market capitalization.

Panel A: Total Sample

	Coeff	t-stat	%sign	Adj R Mean	Adj R Med	Coeff	t-stat	%sign	Adj R Mean	Adj R Med	Coeff	t-stat	%sign	Adj R Mean	Adj R Med
	Slope l. 1					Slope l. 3					Slope l. 5				
1 min.	7.75	8.17	90.81%	1.66%	0.89%	8.90	8.59	88.73%	1.67%	0.85%	8.86	8.53	88.73%	1.64%	0.82%
2 min.	12.53	6.87	90.72%	2.10%	1.36%	13.72	7.25	88.35%	2.01%	1.38%	13.85	7.20	87.97%	1.98%	1.35%
5 min.	20.00	7.79	89.87%	2.77%	1.58%	22.99	5.35	87.69%	2.39%	1.83%	22.32	5.31	87.69%	2.32%	1.77%
10 min.	24.99	5.05	85.80%	2.73%	1.60%	35.48	4.14	86.36%	2.52%	1.99%	34.26	4.10	85.04%	2.43%	1.89%
15 min.	31.48	3.97	80.97%	2.67%	1.56%	38.55	3.50	82.48%	2.59%	1.95%	39.18	3.47	81.63%	2.51%	1.85%
30 min.	34.99	2.80	59.85%	2.28%	1.22%	47.50	2.47	62.97%	2.42%	1.55%	49.51	2.44	61.65%	2.32%	1.51%
	Depth (30 b.p.)					Depth (70 b.p.)					Depth (100 b.p.)				
1 min.	2.25	10.28	94.41%	1.69%	1.59%	2.21	7.76	93.01%	1.01%	0.89%	2.17	6.92	92.45%	0.76%	0.69%
2 min.	3.44	8.69	94.27%	2.15%	1.98%	3.47	6.71	93.43%	1.41%	1.20%	3.42	5.93	91.75%	1.06%	0.96%
5 min.	5.86	6.39	92.86%	2.57%	2.36%	6.21	5.16	91.74%	1.90%	1.71%	6.27	4.57	89.92%	1.46%	1.30%
10 min.	7.66	4.70	90.36%	2.53%	2.32%	8.55	4.03	86.87%	2.13%	1.93%	8.66	3.55	84.78%	1.64%	1.41%
15 min.	8.28	3.84	87.71%	2.44%	2.18%	9.35	3.36	82.68%	2.20%	1.76%	9.52	2.96	77.51%	1.71%	1.41%
30 min.	9.63	2.48	61.79%	1.99%	1.43%	11.12	2.32	60.53%	1.99%	1.43%	11.54	2.08	54.53%	1.62%	1.10%
	CT (m)					CT (m+sd)					CT (m+2sd)				
1 min.	3.26	7.96	92.47%	1.39%	0.91%	2.81	8.15	90.73%	1.35%	0.73%	2.74	7.53	90.50%	1.11%	0.63%
2 min.	5.14	6.94	92.13%	1.81%	1.49%	4.58	7.18	91.09%	1.85%	1.27%	4.51	6.65	90.05%	1.56%	1.06%
5 min.	8.68	5.37	91.66%	2.26%	2.12%	8.12	5.64	90.27%	2.44%	2.07%	8.13	5.28	88.76%	2.14%	1.81%
10 min.	12.40	4.24	90.54%	2.50%	2.33%	11.86	4.53	89.27%	2.77%	2.53%	12.07	4.29	88.12%	2.49%	2.19%
15 min.	14.82	3.66	88.34%	2.68%	2.30%	14.40	3.92	88.34%	2.93%	2.66%	14.46	3.68	85.68%	2.63%	2.34%
30 min.	18.48	2.66	69.32%	2.59%	2.02%	18.62	2.87	75.43%	2.89%	2.31%	18.93	2.71	73.01%	2.64%	2.05%

Table 1.4 continued.

Panel B: Subsamples based on market capitalisation

	Small Cap		Mid Cap		Large Cap		Small Cap		Mid Cap		Large Cap		Small Cap		Mid Cap		Large Cap	
	Adj R^2		Adj R^2		Adj R^2		Adj R^2		Adj R^2		Adj R^2		Adj R^2		Adj R^2		Adj R^2	
	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med	Mean	Med
	Slope l. 1						Slope l. 3						Slope l. 5					
1 min.	0.77%	0.24%	1.33%	0.83%	2.89%	2.22%	0.54%	0.26%	1.27%	0.79%	3.21%	2.75%	0.52%	0.25%	1.24%	0.76%	3.18%	2.66%
2 min.	1.23%	0.40%	2.01%	1.30%	3.06%	2.48%	0.86%	0.45%	1.84%	1.32%	3.34%	3.03%	0.83%	0.44%	1.80%	1.23%	3.30%	2.95%
5 min.	2.17%	0.83%	3.14%	1.71%	2.99%	2.04%	1.60%	0.98%	2.51%	1.95%	3.05%	2.74%	1.52%	0.92%	2.42%	1.84%	3.01%	2.69%
10 min.	2.40%	1.24%	3.29%	1.91%	2.48%	1.59%	2.29%	1.55%	2.82%	2.35%	2.44%	2.04%	2.18%	1.48%	2.69%	2.23%	2.41%	2.04%
15 min.	2.91%	1.64%	3.00%	1.95%	2.08%	1.09%	2.76%	1.97%	2.91%	2.30%	2.09%	1.47%	2.68%	1.91%	2.77%	2.25%	2.06%	1.46%
30 min.	2.83%	1.77%	2.68%	1.55%	1.31%	0.44%	3.16%	2.31%	2.71%	2.02%	1.37%	0.75%	3.10%	2.16%	2.53%	1.90%	1.33%	0.71%
	Depth (30 b.p.)						Depth (70 b.p.)						Depth (100 b.p.)					
1 min.	0.64%	0.35%	1.14%	0.97%	2.44%	2.36%	0.69%	0.47%	0.82%	0.67%	1.26%	1.14%	0.57%	0.37%	0.59%	0.51%	0.95%	0.91%
2 min.	0.93%	0.56%	1.68%	1.50%	2.89%	2.90%	1.05%	0.79%	1.27%	1.10%	1.64%	1.48%	0.87%	0.59%	0.91%	0.79%	1.23%	1.15%
5 min.	1.45%	1.13%	2.31%	2.12%	3.08%	3.08%	1.64%	1.39%	1.93%	1.81%	1.95%	1.74%	1.36%	1.10%	1.46%	1.35%	1.48%	1.32%
10 min.	1.73%	1.39%	2.54%	2.52%	2.75%	2.55%	2.08%	1.90%	2.37%	2.17%	1.93%	1.63%	1.77%	1.59%	1.81%	1.63%	1.47%	1.27%
15 min.	1.92%	1.72%	2.65%	2.41%	2.40%	2.07%	2.47%	2.16%	2.64%	2.31%	1.75%	1.34%	2.06%	1.88%	2.00%	1.81%	1.36%	0.97%
30 min.	2.06%	1.50%	2.33%	1.70%	1.69%	0.95%	2.74%	2.43%	2.54%	2.05%	1.31%	0.63%	2.45%	2.10%	2.00%	1.52%	1.08%	0.53%
	CT (m)						CT (m+sd)						CT (m+2sd)					
1 min.	0.53%	0.37%	1.22%	0.97%	2.62%	2.90%	0.49%	0.33%	1.03%	0.74%	2.76%	3.04%	0.44%	0.28%	0.82%	0.59%	2.26%	2.43%
2 min.	0.85%	0.66%	1.80%	1.64%	2.95%	3.12%	0.80%	0.59%	1.60%	1.24%	3.40%	3.69%	0.72%	0.51%	1.30%	0.98%	2.91%	3.09%
5 min.	1.54%	1.33%	2.49%	2.41%	2.80%	2.85%	1.43%	1.16%	2.39%	2.13%	3.70%	3.73%	1.28%	1.00%	2.00%	1.80%	3.34%	3.29%
10 min.	2.16%	2.03%	2.92%	2.87%	2.34%	2.12%	2.06%	1.88%	2.91%	2.81%	3.41%	3.19%	1.87%	1.64%	2.50%	2.28%	3.19%	3.12%
15 min.	2.69%	2.44%	3.23%	2.87%	1.95%	1.53%	2.55%	2.25%	3.21%	3.06%	3.02%	2.70%	2.29%	1.97%	2.76%	2.53%	2.87%	2.51%
30 min.	3.17%	2.76%	3.04%	2.45%	1.32%	0.71%	3.20%	2.77%	3.19%	2.65%	2.11%	1.45%	2.92%	2.36%	2.81%	2.21%	2.07%	1.49%

creases for longer time horizons. For instance, the mean (median) adjusted R^2 is 0.54% (0.26%) for *Slope* ($l.3$) at a 1 minute horizon, and then increases up to 3.16% (2.31%) at a 30-minute horizon. For large caps we observe a mean (median) adjusted R^2 of 3.21% (2.75%) at a 1 minute horizon and 1.37% (0.75%) at a 30-minute horizon. The pattern is similar for other measures. This opposite pattern suggests that there is a large difference in predictability between stocks of a different market capitalization, which can proxy for a different liquidity or trading activity.[6] For mid-cap stocks we find in-between results.

A faster decrease in predictability over time for larger stocks is in line with the order choice explanation, and not with the informed trading explanation. If predictability is transitory it decreases at a faster rate for more actively traded stocks, as arbitrageurs or informed traders step in and remove predictability at a faster pace. Then why does predictability increase at first, and why does it increase slower for less actively traded stocks? Because it takes an order submission (or cancellation, or modification) to move the price. In our sample the order book is not updated every minute for all stocks. In fact, from Panel C in Table 1.1 we observe that for large cap stocks there are on average (median) 2,905.6 (1,943) updates per day. But even for this subsample in the 5th percentile we observe only 315 updates. For mid cap stocks the average (median) drops to 509.83 (217), and for small caps it is only 131.51 (60), while there are 510 minutes in a trading day. As a consequence, for less actively traded stocks it takes more time for prices to adjust to the state of the book, while it also takes more time for prices of these stock to revert and remove the transitory effect.

1.4.2 Time of Day Patterns in Predictability

We now proceed with our second analysis to address the determinants of return predictability. In particular we study time of day patterns in predictability. To do so we classify all observations for a stock into five subperiods of 102 minutes each (1/5th of the trading day), based on the time of the day and re-estimate Equation (1.1) for each of those five subsamples. By documenting the intraday pattern of return predictability based on book information we aim to gain further insight into the drivers of predictability.

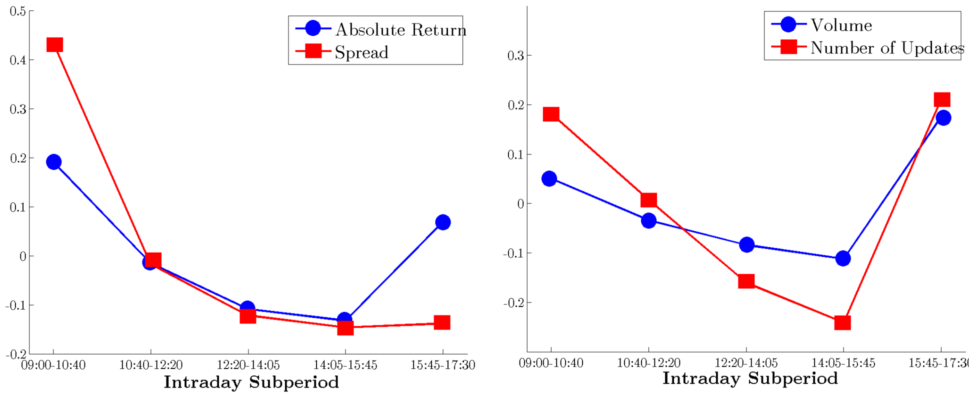
We first document the behavior of a number of market variables throughout the trading day in our sample. Figure 1.1 shows intraday patterns for volatility, the relative bid-ask spread, trading volume and the number of limit order book updates during each of our intraday subperiods, based on 5 minute snapshots of the order book. To be able to compare the intraday pattern across stocks and over time, we standardize these variables by subtracting the stock and day specific mean, and dividing by the standard deviation. In Panel (a) we plot the absolute value of the return, as a proxy for realized volatility over an interval,

[6]We also look at subsamples based on number of trades or volume and find a similar pattern.

Figure 1.1: Intraday Patterns in Market Variables

This figure presents intraday patterns of a number of market variables. For each stock in our sample we measure the market variable at each 5-minute observation and standardize variables by subtracting the day-specific mean and dividing by the standard deviation. These figures show the mean values of these standardized variables at five intraday subperiods of equal length (see Subsection 1.4.2), for the entire sample. Panel (a) shows the intraday pattern of the absolute value of the price return and the quoted bid-ask spread, while Panel (b) focuses on the volume and number of limit order book updates and Panel (c) on the *Depth* (X) measure. *Depth* (X) combines bid and ask depth expressed in thousands of euro and is measured as X basis points away from the midquote.

(a) Volatility and Spread **(b) Trading and Quoting**



and the relative bid-ask spread at the end of the snapshot as a proxy for liquidity.[7] Trading volume and the number of book updates (order book activity) are plotted in Panel (b).

Consistent with previous research (see, e.g., Admati and Pfleiderer, 1988; Garvey and Wu, 2009) we find that these market variables show predictable U-shaped (or reverse J-shaped) patterns throughout the trading day. The order book is the most illiquid in the first period of the trading day when the bid-ask spread is on average more than 0.4 standard deviations above its intraday mean value. Then the market becomes more liquid throughout the day. Volatility is also the highest at the start of the trading day and then decreases gradually throughout the morning period. It is the lowest in the fourth subperiod, after which it rises again. Volume and the number of order book updates follow a similar pattern. The fourth subperiod is thus on average the quietest of the trading day: it has the lowest volatility, the smallest bid-ask spread, lowest volumes and lowest number of book updates. By contrast, trading and quoting activity are concentrated in the first and final subperiod. The first subperiod is also the most illiquid and most volatile of the day.

The (inverted) U-shape of liquidity, volatility and trading and quoting activity can be related to the existence of asymmetric information in financial mar-

[7] A similar pattern is obtained for other liquidity variables, i.e. *Slope*, *Depth* (X) and *CRT*. Figures are available upon request.

Table 1.5: Return predictability throughout the trading day

This table presents and analyzes estimation results of the different models from Equation (1.1) using subsamples based on the time of the day. We distinguish between five periods of equal size throughout the trading day. The independent variable is the return measured during an interval of 5 minutes (expressed in basis points). Explanatory variables are imbalances in the order book state as defined in Subsection 1.2.2. The models are estimated on a per stock and per month basis for the 88 stocks in our sample, for a total of 1,056 stock-months. When an imbalance measure is not defined or only has values that are different from zero for less than 50% of the observations, we do not include the stock-month in our sample for that type of measure. For the slope measure we have results for the full stock-month sample, for the Depth (X) measure we have 715 stock-months (67.71% of the sample), for the CT measure we have 863 stock-months (81.72% of the sample) .In Panel A we present a summary of the results in five columns for different imbalance measures. The first column shows the average coefficient, the second column shows the average t-statistic based on Newey-West HAC standard errors. The third column shows the percentage of regressions for which the coefficient of the imbalance is significantly positive at the 5% level. The fourth column shows the mean adjusted R^2 and the median is shown in the fifth column.

In Panel B we present the mean difference in the adjusted R^2 of the models estimated for different subsamples based on the time of the day. There are 9 subpanels, 1 for each different book imbalance measure we use. Each subpanel contains 5 rows and 5 columns. The $\{k,l\}$-element of such a subpanel is the mean difference in adjusted R^2 between the same model estimated for subperiod k and subperiod l, i.e. for each model j, the mean over all stocks i and months m of $\overline{R}^2_{i,m,j,k} - \overline{R}^2_{i,m,j,l}$. By construction, values below the diagonal are the negative mirror image of values above the diagonal. In superscript we denote how significant the difference in adjusted R^2 is using a t-test, where *, **, *** denotes significance at the 10%, 5% and 1% level respectively. In subscript we use the same notation to present results on the non-parametric Wilcoxon signed-rank test.

Panel A: Summary of Results

Period	Coeff	t-stat	%sign	Adj R^2 Mean	Adj R^2 Med	Coeff	t-stat	%sign	Adj R^2 Mean	Adj R^2 Med	Coeff	t-stat	%sign	Adj R^2 Mean	Adj R^2 Med
	Slope l. 1					Slope l. 3					Slope l. 5				
1	21.72	3.00	63.54%	2.47%	1.45%	34.71	2.60	65.15%	2.44%	1.83%	32.90	2.56	63.54%	2.36%	1.67%
2	19.01	4.77	68.56%	2.89%	1.95%	22.14	2.91	70.17%	2.99%	2.34%	24.00	2.88	69.41%	2.92%	2.27%
3	16.40	2.78	68.37%	2.77%	1.85%	18.59	2.97	71.12%	3.09%	2.24%	19.67	2.95	70.27%	3.05%	2.20%
4	13.63	2.83	60.61%	2.27%	1.45%	14.52	2.63	64.68%	2.54%	1.79%	14.42	2.60	63.83%	2.48%	1.74%
5	20.01	2.55	64.58%	2.37%	1.41%	21.53	2.73	66.57%	2.56%	1.70%	19.91	2.69	65.06%	2.50%	1.64%
	Depth (30 b.p.)					Depth (70 b.p.)					Depth (100 b.p.)				
1	7.31	3.05	78.57%	2.63%	2.14%	8.07	2.72	70.87%	2.45%	2.03%	8.26	2.44	66.53%	1.99%	1.56%
2	6.19	3.34	83.45%	3.21%	2.77%	6.28	2.63	72.83%	2.20%	1.78%	6.41	2.32	63.03%	1.68%	1.27%
3	4.60	3.29	82.89%	3.09%	2.68%	4.38	2.53	70.03%	1.93%	1.57%	4.61	2.27	61.62%	1.47%	1.14%
4	4.65	3.01	78.54%	2.76%	2.23%	4.60	2.30	61.06%	1.73%	1.26%	4.59	2.04	51.68%	1.28%	0.86%
5	7.07	3.25	78.85%	2.98%	2.43%	8.11	2.69	71.01%	2.22%	1.72%	7.89	2.33	59.66%	1.60%	1.22%
	CT (m)					CT (m+sd)					CT (m+2sd)				
1	11.85	2.69	69.69%	2.73%	2.24%	11.34	2.83	71.49%	2.82%	2.33%	11.85	2.68	70.80%	2.54%	1.98%
2	9.05	2.90	74.74%	2.97%	2.58%	8.29	2.97	75.20%	3.04%	2.52%	8.14	2.77	72.31%	2.63%	2.10%
3	7.88	2.84	73.49%	2.82%	2.45%	7.26	2.94	74.04%	2.86%	2.34%	7.06	2.72	69.76%	2.44%	1.95%
4	6.39	2.48	63.75%	2.26%	1.87%	6.44	2.65	65.35%	2.46%	1.71%	6.62	2.46	60.72%	2.10%	1.46%
5	8.88	2.70	70.20%	2.40%	1.98%	8.21	2.97	73.70%	2.87%	2.20%	8.14	2.81	69.29%	2.57%	1.90%

Table 1.5 continued.

Panel B: Mean difference in adjusted R^2 between periods

	Slope l. 1					Slope l. 3					Slope l. 5				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1		$-0.47\%^{**}_{***}$	$-0.23\%_{***}$	0.16%	0.10%		$-0.56\%^{***}_{***}$	$-0.57\%^{***}_{***}$	-0.13%	$-0.15\%^{*}$		$-0.57\%^{***}_{***}$	$-0.62\%^{***}_{***}$	$-0.15\%^{*}$	$-0.18\%^{**}_{*}$
2	$0.47\%^{**}_{***}$		0.24%	$0.63\%^{***}_{***}$	$0.57\%^{***}_{***}$	$0.56\%^{***}_{***}$		-0.02%	$0.43\%^{***}_{***}$	$0.41\%^{***}_{***}$	$0.57\%^{***}_{***}$		-0.04%	$0.43\%^{***}_{***}$	$0.40\%^{***}_{***}$
3	$0.23\%_{***}$	-0.24%		$0.39\%^{***}_{***}$	$0.33\%^{***}_{***}$	$0.57\%^{***}_{***}$	0.02%		$0.45\%^{***}_{***}$	$0.43\%^{***}_{***}$	$0.62\%^{***}_{***}$	0.04%		$0.47\%^{***}_{***}$	$0.44\%^{***}_{***}$
4	-0.16%	$-0.63\%^{***}_{***}$	$-0.39\%^{***}_{***}$		-0.06%	0.13%	$-0.43\%^{***}_{***}$	$-0.45\%^{***}_{***}$		-0.02%	$0.15\%^{*}$	$-0.43\%^{***}_{***}$	$-0.47\%^{***}_{***}$		-0.03%
5	-0.10%	$-0.57\%^{***}_{***}$	$-0.33\%^{***}_{***}$	0.06%		$0.15\%^{*}$	$-0.41\%^{***}_{***}$	$-0.43\%^{***}_{***}$	0.02%		$0.18\%^{**}_{*}$	$-0.40\%^{***}_{***}$	$-0.44\%^{***}_{***}$	0.03%	
	Depth (30 b.p.)					Depth (70 b.p.)					Depth (100 b.p.)				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1		$-0.69\%^{***}_{***}$	$-0.55\%^{***}_{***}$	$-0.23\%^{**}_{*}$	$-0.44\%^{***}_{***}$		0.09%	$0.30\%^{***}_{***}$	$0.59\%^{***}_{***}$	$0.13\%^{*}$		$0.15\%^{**}_{*}$	$0.30\%^{***}_{***}$	$0.60\%^{***}_{***}$	$0.25\%^{***}_{***}$
2	$0.69\%^{***}_{***}$		0.10%	$0.43\%^{***}_{***}$	$0.23\%^{**}_{**}$	-0.09%		$0.21\%^{***}_{***}$	$0.48\%^{***}_{***}$	0.03%	$-0.15\%^{**}_{*}$		$0.14\%^{**}_{**}$	$0.44\%^{***}_{***}$	0.09%
3	$0.55\%^{***}_{***}$	-0.10%		$0.32\%^{***}_{***}$	0.09%	$-0.30\%^{***}_{***}$	$-0.21\%^{***}_{***}$		$0.27\%^{***}_{***}$	$-0.18\%^{***}_{***}$	$-0.30\%^{***}_{***}$	$-0.14\%^{**}_{**}$		$0.30\%^{***}_{***}$	-0.05%
4	$0.23\%^{**}_{*}$	$-0.43\%^{***}_{***}$	$-0.32\%^{***}_{***}$		$-0.22\%^{**}_{**}$	$-0.59\%^{***}_{***}$	$-0.48\%^{***}_{***}$	$-0.27\%^{***}_{***}$		$-0.45\%^{***}_{***}$	$-0.60\%^{***}_{***}$	$-0.44\%^{***}_{***}$	$-0.30\%^{***}_{***}$		$-0.35\%^{***}_{***}$
5	$0.44\%^{***}_{***}$	$-0.23\%^{**}_{**}$	-0.09%	$0.22\%^{**}_{**}$		$-0.13\%^{*}$	-0.03%	$0.18\%^{***}_{***}$	$0.45\%^{***}_{***}$		$-0.25\%^{***}_{***}$	-0.09%	0.05%	$0.35\%^{***}_{***}$	
	CT (m)					CT (m+sd)					CT (m+2sd)				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1		$-0.26\%^{***}_{***}$	$-0.11\%_{*}$	$0.40\%^{***}_{***}$	$0.27\%^{***}_{**}$		$-0.32\%^{***}_{***}$	$-0.23\%^{***}_{**}$	$0.17\%^{**}_{***}$	-0.12%		$-0.26\%^{***}_{***}$	$-0.18\%^{**}$	$0.22\%^{***}_{***}$	$-0.14\%^{**}$
2	$0.26\%^{***}_{***}$		0.12%	$0.60\%^{***}_{***}$	$0.48\%^{***}_{***}$	$0.32\%^{***}_{***}$		0.09%	$0.49\%^{***}_{***}$	$0.20\%^{**}_{**}$	$0.26\%^{***}_{***}$		0.08%	$0.47\%^{***}_{***}$	$0.11\%^{*}$
3	$0.11\%_{*}$	-0.12%		$0.49\%^{***}_{***}$	$0.36\%^{***}_{***}$	$0.23\%^{***}_{**}$	-0.09%		$0.41\%^{***}_{***}$	0.12%	$0.18\%^{**}$	-0.08%		$0.40\%^{***}_{***}$	0.04%
4	$-0.40\%^{***}_{***}$	$-0.60\%^{***}_{***}$	$-0.49\%^{***}_{***}$		$-0.14\%^{*}_{**}$	$-0.17\%^{**}_{***}$	$-0.49\%^{***}_{***}$	$-0.41\%^{***}_{***}$		$-0.29\%^{***}_{***}$	$-0.22\%^{***}_{***}$	$-0.47\%^{***}_{***}$	$-0.40\%^{***}_{***}$		$-0.36\%^{***}_{***}$
5	$-0.27\%^{***}_{**}$	$-0.48\%^{***}_{***}$	$-0.36\%^{***}_{***}$	$0.14\%^{*}_{**}$		0.12%	$-0.20\%^{**}_{**}$	-0.12%	$0.29\%^{***}_{***}$		$0.14\%^{*}$	$-0.11\%_{*}$	-0.04%	$0.36\%^{***}_{***}$	

kets and the behavior of informed traders. Asymmetric information is generally higher at the start of the trading day, when new information that has arrived overnight still needs to be processed by market participants. During this period the mispricing is the largest, creating an opportunity for higher trading profits to informed traders. Kaniel and Liu (2006) and Roşu (2015) show theoretically that informed traders are more likely to submit market orders under these circumstances. Bloomfield, O'Hara, and Saar (2005) confirm in an experimental setting that informed traders indeed submit more market orders at the start of the trading day, and switch to limit orders later. If predictability is caused by informed traders who submit limit orders, then return predictability should increase during the trading day. Because informed traders may switch to market orders also at the end of the trading day, when they hide among liquidity traders, predictability should also be lower at the end of the trading day if the informed trading hypothesis holds. Because it is unclear which intraday pattern is to be expected if predictability is caused by traders endogenously adjusting their order choice, we consider the analysis in this Subsection mainly as a test of the informed trading hypothesis.

Table 1.5 Panel A shows a summary of estimation results for Equation (1.1), estimated using the same nine imbalance measures as before, for each of the five subsamples based on the time of the day. We use 5 minute returns and the table representation is similar as before. Panel B examines more closely the differences in adjusted R^2 between the different subperiods. In each row k we show the mean difference in adjusted R^2 between the model estimated for subperiod k, and the other four subperiods (in one of five columns). A positive (negative) difference indicates that the adjusted R^2 for subperiod k is on average larger (smaller) than the adjusted R^2 for the subperiod from column l. In superscript we denote whether this difference is significant according to a paired t-test. In addition, we also apply the non-parametric Wilcoxon signed rank test, of which significance is indicated in subscript. For instance, in row 2, column 4 for the $Slope$ $(l.1)$ measure, we find that the adjusted R^2 is on average 0.63% higher in the second subperiod, as compared to the fourth subperiod. Both the t-test and Wilcoxon signed rank test indicate that this difference is significant.

In general we find that predictability is the highest during the second and third subperiod of the trading day, and it is the lowest during the fourth subperiod, which corresponds to the quietest period of the day. This could be due to the fact that returns are less predictable when there is on average less trading and quoting activity during a time interval, such as for small caps at the shortest time horizons of Table 1.4 Panel B. Similarly, during the relatively calm fourth subperiod of the trading day there is less trading and quoting activity within any 5 minute interval, which weakens the relation between the lagged state of the order book and returns.

Interestingly, both the first and final subperiod of the trading day have intermediate predictability. This is not in line with the predicted pattern according to the informed trading hypothesis. Furthermore, it is remarkable that there is no (inverted) U-shape in predictability, as is the case for most market variables, and that the largest differences in predictability are between the largely similar middle subperiods. This suggests that predictability is not monotonically related to a single market variable. In any case, multiple drivers of predictability are likely to be a work, which leads to an ambiguous intraday pattern. For instance, while liquidity traders condition their order choice on the state of the book, the use of market orders by informed traders could remove transient price effects that are the result of the former behavior, the more so at the start and end of the trading day.

1.4.3 Cross-sectional Determinants of Predictability

In this subsection we turn to the third analysis to investigate the drivers of predictability. In particular, we are most interested in whether predictability is related to informed trading. To measure informed trading we estimate the probability of informed trading (PIN) developed by Easley, Kiefer, and O'Hara (1996) (see Appendix A for details). The PIN is an estimate of the probability that on a given day within the estimation period an order comes from an informed trader. We estimate the PIN for each stock and each month in the sample. The optimization converges for 1,001 of 1,056 stock-months. The average $PIN_{i,m}$ for all stocks i and months m is 22.21%, ranging between 9.84% in the 5th percentile and 42.95% in the 95th percentile (Table 1.1 Panel A). These values are in line with previous literature (see, e.g., Easley, Hvidkjaer, and O'Hara, 2002; Duarte and Young, 2009). From Panel B of Table 1.1 it is also clear that the PIN is the largest for small cap stocks and smallest for large cap stocks, which is in line with economic theory. If the informed trading explanation for return probability is valid, we expect a positive relation between the PIN and our measures of return predictability.

Since we obtain monthly results for our predictability and informed trading measure we opt for a panel data analysis using the Fama-MacBeth approach. This allows us to exploit the cross-sectional variation in predictability and informed trading. Our main question is for which type of stocks the predictive ability of order book information for future returns is larger. More specifically, we estimate monthly cross-sectional regressions of $\overline{R}^2_{i,m,j}$, the adjusted R^2 of a model using order book measure j of stock i for month m, on a number of stock characteristics. We use the adjusted R^2 obtained from estimation on 5 and 30 minute intervals. In order to account for autocorrelation in the error term we use Newey-West adjusted standard errors. We estimate two specifications. In the first one we control for the natural logarithm of the average market value $Ln(MV_{i,m})$, which correlates with a number of market variables that could impact return predictability. The second specification replaces the market value by several stock characteristics: the natural logarithm of average euro depth quoted at the best prices $Ln(Depth_{i,m})$, the natural logarithm of the average daily euro trading volume $Ln(Volume_{i,m})$, the standard deviation of daily returns $Volat_{i,m}$ and the average daily relative bid-ask spread $Spread_{i,m}$. The models to be estimated are then as follows (with $\zeta_{i,m}$ and $\eta_{i,m}$ the error terms):

$$\overline{R}^2_{i,m,j} = \beta_0 + \beta_1 PIN_{i,m} + \beta_2 Ln(MV_{i,m}) + \zeta_{i,m}, \tag{1.9}$$

$$\begin{aligned}\overline{R}^2_{i,m,j} = \; & \beta_0 + \beta_1 PIN_{i,m} + \beta_2 Ln(Depth)_{i,m} \\ & + \beta_3 Ln(Volume_{i,m}) + \beta_4 Volat_{i,m} + \beta_5 Spread_{i,m} + \eta_{i,m}.\end{aligned} \tag{1.10}$$

In particular, we are also interested in the effect of quoted depth on return predictability. If the order choice hypothesis holds, then we expect returns to be more predictable for stocks with deeper order books. Traders arriving to the market are more likely to adjust their order choice when they observe a thick order book on their side, e.g., by switching from a limit order at the back of the queue to a price improving limit order or a market order. Indeed, deep order books are indicative of a competitive liquidity supply. These are precisely the circumstances that cause short-term returns to be predictable according to the order choice hypothesis.

The estimation results of Equations (1.9) and (1.10) are presented in Table 1.6. Panels A and B show results for the adjusted R^2 from the *Slope* measure models, for 5 minute returns and 30 minute returns respectively, Panels C and D for the *Depth* (X) measure, and Panels E and F for the CT measure. The first two lines of each panel present coefficients and t-statistics for the PIN, our measure of informed trading. In general, informed trading is negatively or insignificantly related to return predictability based on order book information (with one exception in Panel F). This is a surprising result, as it contradicts the hypothesis that return predictability is caused by informed traders that submit informative limit orders to the book. Furthermore, we find in 16 out of 18 models that return predictability is significantly higher for stocks with deeper order books, in line with the order choice hypothesis. Return predictability is thus transitory in nature and is caused by traders adjusting their order choice as a consequence of observed limit order book imbalances. The presence of more informed traders for some stocks may actually help to reduce return predictability, as indicated by the negative coefficients. Informed traders trade in the opposite direction of the transitory price movement and drive prices back to the efficient price.

Furthermore, for 30 minute returns predictability is lower for stocks with a larger market capitalization and higher trading volume, in line with Panel B Table 1.4. When more traders are actively trading a stock, in the long run their actions help to eliminate any mispricing that is the consequence of transient price effects in the limit order book. For 5 minute returns, the effect of volume

Table 1.6: Determinants of predictability

This table presents estimates of Equations (1.9) and (1.10). The models are estimated using the Fama-MacBeth procedure using monthly regressions. The dependent variable is the adjusted R^2 from the return prediction models using order book imbalances. Results are presented for regressions using different order book imbalances: $Slope$ (Panels A and B), $Depth\ (X)$ (Panels C and D) and CT (Panels E and F), and different time horizons (5 minutes versus 30 minutes). The independent variables are the probability of informed trading PIN (see), the natural logarithm of market value, the natural logarithm of the daily average euro depth at the best quotes (level 1), the natural logarithm of average daily euro volume, the standard deviation of daily returns and the daily average relative bid ask-spread (time-weighted throughout the trading day) (time-weighted throughout the trading day). All variables are winsorized at the 1% and 99% level. t-statistics, adjusted for serial correlation using the Newey-West procedure, are presented in parentheses below the coefficient estimates. *, **, *** denote significance at the 10%, 5% and 1% level respectively.

Panel A: Slope at 5 minute intervals

	1.1		1.3		1.5	
PIN	-0.016*	0.006	-0.026***	-0.015***	-0.005	-0.018***
	(-1.90)	(0.61)	(-5.97)	(-3.56)	(-1.02)	(-4.32)
Ln(MV)	0.002***		0.002***		-0.005***	
	(3.99)		(9.04)		(-20.76)	
Ln(Depth)		0.013***		0.012***		0.013***
		(3.13)		(6.03)		(6.70)
Ln(Volume)		-0.005**		-0.000		-0.000
		(-2.79)		(-0.26)		(-0.40)
Volat		1.455***		0.529***		0.531***
		(5.82)		(4.48)		(4.79)
Spread		-1.785***		0.009		0.009
		(-4.30)		(0.04)		(0.04)
Const	0.020***	-0.065	0.015***	-0.115***	0.059***	-0.118***
	(4.56)	(-1.78)	(8.76)	(-10.96)	(23.03)	(-12.15)
Obs	989	989	989	989	989	989
R^2	0.018	0.139	0.064	0.242	0.118	0.247

Panel B: Slope at 30 minute intervals

	1.1		1.3		1.5	
PIN	-0.004	-0.018	-0.005	-0.023**	-0.005	-0.023**
	(-0.52)	(-1.29)	(-0.87)	(-3.02)	(-1.02)	(-3.09)
Ln(MV)	-0.005***		-0.005***		-0.005***	
	(-12.29)		(-16.62)		(-20.76)	
Ln(Depth)		0.014***		0.010***		0.010***
		(5.25)		(3.58)		(4.08)
Ln(Volume)		-0.008***		-0.006***		-0.006***
		(-5.42)		(-4.65)		(-4.92)
Volat		0.334		0.015		-0.012
		(1.37)		(0.09)		(-0.10)
Spread		0.455		0.723		0.641
		(0.63)		(1.39)		(1.39)
Const	0.055***	-0.025	0.060***	0.001	0.059***	0.001
	(14.68)	(-1.08)	(20.94)	(0.07)	(23.03)	(0.09)
Obs	989	989	989	989	989	989
R^2	0.092	0.193	0.111	0.184	0.118	0.173

Table 1.6 continued.

Panel C: Depth(X) at 5 minute intervals

	(30 b.p.)		(70 b.p.)		(100 b.p.)	
PIN	-0.017***	-0.006	-0.031***	-0.027***	-0.026***	-0.025***
	(-3.51)	(-0.92)	(-6.25)	(-3.82)	(-7.78)	(-5.14)
Ln(MV)	0.005***		0.001		-0.000	
	(9.57)		(1.01)		(-1.31)	
Ln(Depth)		0.004		0.009***		0.007***
		(1.71)		(5.82)		(5.43)
Ln(Volume)		0.002		-0.000		-0.000
		(1.55)		(-0.66)		(-0.21)
Volat		0.080		0.290**		0.274**
		(0.47)		(2.53)		(2.30)
Spread		-1.849***		0.616*		0.859***
		(-6.41)		(2.12)		(3.31)
Const	-0.006	-0.037**	0.022***	-0.079***	0.025***	-0.062***
	(-1.34)	(-2.54)	(5.86)	(-6.75)	(9.78)	(-8.39)
Obs	704	704	943	943	974	974
R^2	0.168	0.294	0.060	0.212	0.054	0.194

Panel D: Depth(X) at 30 minute intervals

	(30 b.p.)		(70 b.p.)		(100 b.p.)	
PIN	0.019	-0.001	-0.014*	-0.037***	-0.011**	-0.036***
	(1.40)	(-0.10)	(-1.93)	(-3.62)	(-2.29)	(-5.37)
Ln(MV)	-0.001**		-0.004***		-0.005***	
	(-2.95)		(-8.25)		(-11.85)	
Ln(Depth)		0.009***		0.009***		0.008***
		(3.49)		(7.36)		(7.08)
Ln(Volume)		-0.006***		-0.005***		-0.004***
		(-4.60)		(-6.81)		(-5.78)
Volat		-0.530***		-0.486***		-0.322***
		(-3.72)		(-5.22)		(-3.49)
Spread		-1.623*		1.402***		1.965***
		(-1.88)		(3.11)		(3.53)
Const	0.024***	0.021	0.056***	-0.007	0.061***	-0.009
	(4.47)	(1.66)	(10.97)	(-0.82)	(17.52)	(-1.27)
Obs	704	704	943	943	974	974
R^2	0.050	0.193	0.090	0.173	0.144	0.215

Table 1.6 continued.

Panel E: CT at 5 minute intervals

	(m)		(m+sd)		(m+2sd)	
PIN	-0.013**	0.005	-0.034***	-0.018**	-0.035***	-0.021***
	(-2.70)	(1.17)	(-5.65)	(-3.10)	(-6.74)	(-4.34)
Ln(MV)	0.003***		0.004***		0.004***	
	(12.64)		(21.30)		(16.26)	
Ln(Depth)		-0.002		0.005**		0.010***
		(-1.33)		(2.24)		(4.66)
Ln(Volume)		0.003***		0.003**		0.001
		(4.13)		(3.02)		(1.05)
Volat		-0.075		0.108		0.360**
		(-0.84)		(0.98)		(2.59)
Spread		-1.006**		-0.344*		-0.214
		(-2.98)		(-1.94)		(-0.95)
Const	0.004**	0.010	0.006***	-0.063***	0.009***	-0.098***
	(2.35)	(1.11)	(4.94)	(-4.58)	(4.83)	(-7.07)
Obs	878	878	964	964	980	980
R^2	0.138	0.277	0.177	0.297	0.142	0.291

Panel F: CT at 30 minute intervals

	(m)		(m+sd)		(m+2sd)	
PIN	0.022***	-0.001	-0.004	-0.027***	-0.005	-0.029***
	(4.99)	(-0.10)	(-0.76)	(-3.62)	(-1.11)	(-4.09)
Ln(MV)	-0.004***		-0.004***		-0.004***	
	(-8.86)		(-13.26)		(-9.19)	
Ln(Depth)		0.009***		0.009***		0.013***
		(3.49)		(4.04)		(6.13)
Ln(Volume)		-0.006***		-0.008***		-0.008***
		(-4.60)		(-7.10)		(-6.31)
Volat		-0.530***		-0.471***		-0.345**
		(-3.72)		(-3.38)		(-2.23)
Spread		-1.623*		0.037		0.214
		(-1.88)		(0.12)		(0.51)
Const	0.048***	0.021	0.061***	0.041**	0.059***	0.007
	(13.28)	(1.66)	(16.57)	(2.98)	(12.25)	(0.84)
Obs	878	878	964	964	980	980
R^2	0.103	0.193	0.088	0.186	0.085	0.186

is not consistent across specifications. At this frequency trading volume could be either increasing or decreasing predictability. Predictability is positively related to volatility for 5 minute returns, but negatively for 30 minute returns. Returns over longer intervals are inherently more difficult to predict for more volatile stocks, but for shorter intervals the order book actually contributes in explaining returns. Lower volatility stocks may exhibit more zero return intervals over shorter time periods, which causes returns to be less predictable.

1.5 Conclusion

In this chapter we investigate the relation between the state of the order book as displayed on a trader's screen and future returns. We first show that order book information can indeed be used to predict returns in the short-term. Furthermore, measures that are confined to the information at the top of the book (i.e. at or close to the best prices in the book) appear to be the best predictors. As more information from deeper in the book is captured in a measure, the predictive ability of the measure decreases. This suggests that the most relevant information for future returns is concentrated at the top of the book.

A more fundamental question is what drives return predictability on the basis of the book. We argue that two potential sources exist. On the one hand, predictability in returns can be due to traders who condition their order choice on the state of the order book. This creates systematic patterns in order flow as patient traders are crowded out the order book. This can generate short-run ('transitory') price effects. We call this the order choice hypothesis. On the other hand, the presence of informed traders who submit limit orders is another potential source of predictability (with 'permanent' price effects). As such, the limit order book should reflect the private information and thus be informative about future returns. This explanation is termed the informed trading hypothesis.

Our results indicate that return predictability is most likely caused by the former effect, while evidence in favor of the informed trading hypothesis is weak. First, looking at predictability at short versus longer horizons, return predictability based on book information increases from a 1 minute until a 15 minute horizon. But at horizons between 15 and 30 minutes, return predictability declines again, while it does not completely vanish. As such, we find that there is both a transitory component and a permanent component to return predictability. But there are remarkable differences between large cap and small cap stocks. For small cap stocks predictability is the lowest at short horizons, and then decreases gradually for longer horizons. For large cap stocks we observe the opposite pattern. This suggests that the bulk of predictability is transitory, and that predictability decreases much faster for actively traded stocks. Second, we find no obvious and consistent intraday pattern for predictability. When we divide the trading day into five subperiods, predictability is in general the largest for the second and third subperiod, while it is the lowest for the fourth subperiod. Predictability is intermediate at the start and end of the trading day. While the expected intraday pattern of predictability according to the order choice hypothesis is unclear, we would expect to find the lowest predictability at the start and end of the trading if predictability is caused by informed limit orders. Third, we find cross-sectionally that predictability is negatively related to the probability of informed trading, while for stocks with deeper (and thus more competitive) order books predictability is higher. Furthermore, predictability of returns over longer time horizons is much lower for more actively traded stocks. In sum, we show that returns are predictable because the order choice of traders predictably depends on the state of the order book. The effect of informed traders in the limit order book is limited. By contrast, informed traders help to decrease predictability by trading in the opposite direction of the transitory price effect.

The Spanish stock market in 2003 presents a setting in which stocks are traded in a single order book and, to our knowledge, orders are submitted mostly by humans. This allows for a clean test of how limit order book theories determine the relation between the state of the order book and future prices, without considerations of competition between order books and the interference of (high-frequency) algorithmic trading. But during the last decade the market environment has changed. Motivated by technological innovations and spurred by new regulation the trading landscape has fragmented. Stocks around the world are now traded simultaneously on multiple trading venues that compete by offering different matching mechanisms, order types and transparency. Not all traders have access to alternative trading venues, while others employ algorithms, known as smart order routers, to split their orders over trading venues automatically – using the observed state of the order books on these venues as inputs (Foucault and Menkveld, 2008). The latter complicates the interaction between the order book and prices on a single venue, but can also create new interactions between prices and order books of different venues.

Furthermore, trading venues are now populated not only by human traders. A significant portion of orders (and importantly, cancellations) are now submitted by algorithmic traders. By design these algorithms condition their actions on the current and previous state of the market. An important question is then whether the presence of algorithmic traders increases predictability on the basis of the order book by formalizing the order choice considerations that lead to transitory price effects into algorithms, or whether algorithms are able to better predict transitory price effects and trade in the opposite direction, thereby reducing predictability. Brogaard, Hendershott, and Riordan (2014) show that algorithmic traders may engage in the latter type of strategies. Another important consideration for today's markets is that one class of algorithmic traders, high-frequency traders, has made competition on speed its core business. High-frequency traders have taken order book interactions to a millisecond environment. Therefore, if return predictability persists in today's markets, it is likely to be found at these higher frequencies. We leave these questions for future research.

Appendix A: The Probability of Informed Trading

The probability of informed trading or PIN is a concept that originates from microstructure trading models (see, e.g., Glosten and Milgrom, 1985; Easley, Kiefer, and O'Hara, 1996; Easley, Kiefer, O'Hara, and Paperman, 1996; Easley, Hvidkjaer, and O'Hara, 2002) in which traders sequentially arrive in a dealer market. Each trader is potentially better informed about the fundamental value of the security than the dealer (or liquidity supplier). The dealer cannot distinguish informed traders from uninformed traders who trade for liquidity reasons and must set his quotes so as to compensate him for the expected loss of trading against informed traders. Informed traders only enter the market on days when information events occur which is determined by nature on the beginning of each trading day with probability α. Upon occurrence, the probability that the information is bad news is δ, the probability that it is good news is $1-\delta$. When an information event occurs informed traders arrive in the market according to an independent Poisson process with arrival rate μ. Uninformed buyers and sellers arrive with rates ϵ_B and ϵ_S respectively. The probability on any day that an order is from an informed trader is then given by

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s}. \tag{1.11}$$

The parameters of the model can be estimated via maximum likelihood. The likelihood function is

$$\begin{aligned} L(\theta|B,S) \quad = \quad & (1-\alpha)e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!}. \\ & +\alpha\delta e^{-\varepsilon_b}\frac{\varepsilon_b^B}{B!}e^{-(\mu+\varepsilon_s)}\frac{(\mu+\varepsilon_s)^S}{S!} \\ & +\alpha(1-\delta)e^{-\varepsilon_s}\frac{\varepsilon_s^S}{S!}e^{-(\mu+\varepsilon_b)}\frac{(\mu+\varepsilon_b)B}{B!} \end{aligned} \tag{1.12}$$

with B the observed number of buy orders, S the observed number of sell orders and θ the parameters. Given the independence of the processes across trading day, the likelihood function over N trading days is then

$$L(\theta|M) = \prod_{d=1}^{N} L(\theta|B_d, S_d) \tag{1.13}$$

with $(B_d,\ S_d)$ the number of buyer-initiated and seller-initiated orders for day d and $M = ((B_1,\ S_1), ..., (B_N,\ S_N))$. Because estimating these parameters is computationally burdensome, and to increase the probability of convergence we follow Aktas, De Bodt, Declerck, and Van Oppens (2007) and simplify the log-likelihood function to

$$\begin{aligned} Log(L(\theta|M)) \quad = \quad & \sum_{d=1}^{N} \{-2\varepsilon + M_d \ln(x) + (B_d + S_d) \ln(\mu + \varepsilon)\} \\ & + \sum_{d=1}^{N} \ln \left\{ \alpha(1-\alpha) e^{-\mu} x^{S_d - M_d} + \alpha\delta e^{-\mu} x^{B_d - M_d} + (1-\alpha) x^{B_d + S_d - M_d} \right\} \end{aligned} \tag{1.14}$$

with $M_d = \min(B_d, S_d) + \frac{\max(B_d, S_d)}{2}$ and $x = \frac{\varepsilon}{\varepsilon + \mu}$. Hereby we assume that the arrival rates of uninformed buyers and sellers are equal: $\varepsilon_B = \varepsilon_S = \varepsilon$. We maximize the likelihood function over all trading days in a month Following Duarte and Young (2009), we run the optimization five times, with randomly generated different starting points for the parameters to avoid arriving at a local maximum and then select the maximum of the five optimizations.

Chapter 2

Limit Order Book Information and Return Predictability in a Fragmented Market

2.1 Introduction

Imbalances between the bid and ask side of the limit order book have been shown empirically to predict short-term intraday returns. Intuitively, bid and ask liquidity capture unexecuted trading desires of patient buyers and sellers. Previous studies concerning return predictability from the order book state (i.e., publicly observable information on prices and depth), however, focus on concentrated markets, while it is a salient feature of present day financial markets that order flow is fragmented across trading venues. Stocks are now traded not only on the listing exchange, but also on a number of alternative trading venues.[1] This chapter extends research on intraday return predictability by including order book information from multiple trading venues. We examine whether the state of an order book can predict returns on the same, but also on other venues. In line with previous research, we capture the state of an order book by the imbalance

[1]In the U.S. these are known as Electronic Communication Networks (ECNs), while in Europe competition emerges in the form of Multilateral Trading Facilities (MTFs). Both types of venues operate in a similar way: they largely employ transparent and fully automated order books that match third party order flow. Next to these so-called *lit* trading venues there is also competition from *opaque* or *dark* venues, such as dark pools. Furthermore, although limit order books of most trading venues are essentially transparent, traders generally have the option to reduce their order exposure by submitting (partially) hidden orders. De Winne and D'Hondt (2007) show that these orders can provide a substantial part of liquidity in limit order books.

between *visible* bid side depth and ask side depth.[2] We also take into account that the relative position of order books on the price grid can impact their predictive ability, and we document how prices across trading venues adjust to one another during the trading day. Finally, we investigate how the predictive ability of trading venues is affected by fragmentation of trading volume.

Using a sample of 30 stocks listed on Euronext Amsterdam, Brussels or Paris over a one year period (October 2009 - September 2010), we show that information obtained from the order book can predict returns within and across venues. Both imbalances between the bid and ask side, as well as the relative position of the quotes on the price grid matter. We explicitly investigate how the relative quote position can predict future returns by looking at the distance between midquotes of different order books. We find that prices across trading venues have a tendency to adjust to each other. In addition, we show that order book imbalances from other venues are the strongest predictors of future returns when the prices on these venues are more competitive (i.e. a higher bid price and a lower ask price). Order book imbalances obtained from worse prices actually predict reversals (as they have a negative relation with future returns). Our results suggest that some venues are leading in price discovery, while others are followers. Returns on the latter venues are predictable mainly because they adjust to prices of the leading venues. Moreover, their order book has almost no predictive ability for returns on other venues.. By contrast, returns on leading venues are relatively less predictable, and both the imbalance in the state of the book and its position on the price grid have predictive power for returns on other venues.

Predicting returns (and other market variables) at high frequencies is important for the many market participants who rely on algorithms to execute their trades. Algorithms use market data as an input to generate trading decisions.[3] In particular, for traders employing a market making strategy it is important to assess the direction in which prices are likely to move. Managing their risky inventories and assessing the information content of the order flow is at the core of their business. Our results imply that optimally these traders should use information from multiple trading venues. Moreover, it is likely that those traders who are active on multiple trading venues, such as cross-venue market makers, arbitrageurs or brokers using smart order routing technology, play a key role in generating cross-venue return predictability.

There are two potential explanations as to why the state of the order book can predict returns. First, the order book may contain fundamental information

[2] We have no information about the hidden (unexecuted) liquidity in the order book. Market participants can also only observe this part of the order book, although they can estimate whether there is hidden liquidity present (and to which extent). Doing so would, however, take us beyond the scope of this research.

[3] These market data appear to provide useful information for traders, as market data sales are a growing source of revenues for trading venues (Easley, O'Hara, and Yang, 2013; Cespa and Foucault, 2014).

that is impounded into prices with a lag. Second, and quite the reverse, the state of the order book causes temporary price deviations as traders adjust their behavior to the state of the order book. This is because the state of the order book influences the execution probability of their orders. They choose more (less) aggressive orders when the book is thicker (thinner) on their own side of the order book (with reverse predictions for the opposite side of the book, see, e.g., Parlour, 1998; Ranaldo, 2004; Goettler, Parlour, and Rajan, 2005). This explanation implies that the order book predicts returns because it is transparent. Tombeur and Wuyts (2015) find that the latter is also most likely to be the driver of return predictability on the basis of the limit order book.

Both explanations for return predictability also suggest that the order book on one venue can be used to predict returns on *another* venue. If the order book contains fundamental information (e.g., because informed traders submit limit orders on one side of the order book only), order book imbalances that are able to predict returns on their own venue, should also be able to predict returns on other venues. All venues are trading the same asset that is subject to the same information shocks and signals. In efficient markets information that affects prices on one trading venue should also affect prices on other trading venues. Market makers that are active across multiple venues should adjust their quotes on all these venues similarly. If not, then arbitrageurs can take advantage of this, and prices adjust following their actions (Foucault, Kozhan, and Tham, 2015).

In the case that predictability is driven by order choice, when traders are considering multiple venues for their order submissions, it is the state of *all relevant* order books that matters for their choice, and thus can predict returns. Relevant order books are those that a trader considers for order submission himself, and also those that are used by other traders; order execution probability depends on his own choice and choices made by future traders. We show that returns on the listing exchange (which has by far the largest amount of trading volume) are more predictable using order book imbalances from alternative venues when the latter have a larger market share. When market share is a measure of how relevant a trading venue is, this is in line with predictability being driven by endogenous order choice. Traders condition their order choice on the state of the order book of an alternative venue relatively more when the state of its order book matters. However, we find no evidence of returns on alternative venues being more predictable from the order book state of the listing exchange during these times of increases fragmentation, possibly because the state of the order book on the listing exchange is *always* relevant.

Our research contributes to previous work that studies whether the state of the book is informative on future price changes. Harris and Panchapagesan (2005) find that the order book on the New York Stock Exchange contains information on future prices, and that the specialist uses his privileged knowledge of the order book in his favor. Cao, Hansch, and Wang (2009) show that the top of the order book contributes the most to return predictability. Jain, Jain, and McInish (2011) show that the state of the order book can predict trade price movements as well as future volatility. Yamamoto (2012) finds that imbalances in the state of the order book can predict returns, but technical trading strategies based on imbalances in the state of the limit order book cannot outperform a buy-and-hold strategy. Brogaard, Hendershott, and Riordan (2014) use limit order book imbalances as a proxy of short-term information that high-frequency traders (HFTs) can use in their strategies. They show that, following an imbalance, HFTs supplying liquidity submit orders on the thin side of the order book, and thereby help to reduce predictability. But HFT liquidity demanders trade in the direction of the order book imbalances, thus contributing to return predictability. Overall, the latter effect dominates, and HFTs as a group trade in the direction of the order book imbalance. Tombeur and Wuyts (2015) find evidence that the short-term return predictability phenomenon is caused by traders who endogenously choose their order aggressiveness in response to the state of the limit order book. This translates into predictable, but transient, midquote changes.

The remainder of the chapter is organized as follows. Section 2.2 discusses related literature and presents our main hypotheses. The method we use to test these hypotheses is presented in Section 2.3, while Section 2.4 describes the data. Next, we discuss results of our predictability models in Section 2.5. Finally, Section 2.6 concludes.

2.2 Literature and Hypotheses

2.2.1 Return Predictability on the Own Venue

Based on theory models two main reasons have been put forward to explain why imbalances can predict short-term returns: (1) traders condition their order choice on the state of the limit order book and this leads to predictable patterns in the order book, (2) imbalances in the limit order book are related to fundamental information, potentially because informed traders submit limit orders.

For the first explanation, Parlour (1998) shows in a dynamic model for a single limit order book that arriving traders choose their order aggressiveness endogenously, based on the current state of the limit order book. When the book is deeper (thinner) on the own (opposite) side of the market, this decreases (increases) the execution probability of limit orders that are submitted at the same price, as these are executed on the basis of time priority. This generates predictable patterns in order choice. But since the original model features the restrictive assumptions of a one tick spread and fixed bid and ask quotes, it does not allow to make any inference about patterns in quotes. In a similar setting, but with an extended discrete price grid around the common asset value, Goettler, Parlour, and Rajan (2005) show numerically that when traders compete for order execution also by choosing a limit price, they can submit more aggressive limit orders to jump the queue and increase their execution probability. They find that because of such behavior the midquote can deviate from the common value. This is in line with the state of the order book predicting midquote returns because of order choice.

Empirically, Ranaldo (2004) shows that traders submit more (less) aggressive orders when depth on the own (opposite) side of the market is larger (smaller), consistent with traders using the state of the order book in order to asses their execution probability.[4] When, as a result of their increased aggressiveness, traders improve the best prices on their side of the market, or consume the full liquidity available on the other side, this moves the midquote price (i.e. a non-zero return). This explanation of return predictability implies that the price effect of the limit order book state is transient. Only short horizon returns are predictable. After some time other traders arrive and submit orders that drive prices back in line with fundamentals.

The second explanation, fundamental information contained in the limit order book, is usually related to the question of what kind of orders informed traders use: do they demand liquidity or supply liquidity? Theory models argue that under certain conditions informed traders tend to submit limit orders,[5] but empirically it remains a challenging task to find evidence in favor of this hypothesis.[6] If informed traders do submit limit orders they will tend to submit orders on one side of the book only (when the security is undervalued by the current

[4]The finding that the state of the order book matters in explaining order choice also appears to be robust across markets and over time. Similar results are for instance reported by Biais, Hillion, and Spatt (1995), Griffiths, Smith, Turnbull, and White (2000), Beber and Caglio (2005), Cao, Hansch, and Wang (2008), Duong, Kalev, and Krishnamurti (2009) and Pascual and Veredas (2009b).

[5]Traditional models of limit order markets assume that informed traders only demand liquidity while uninformed traders provide liquidity and bear adverse selection costs, similar to a dealer market (e.g., Glosten, 1994; Rock, 1996; Seppi, 1997). There are fewer models that allow informed traders to endogenously choose between market orders and limit orders. Chakravarty and Holden (1995) show that informed traders may use a combination of market and limit orders of opposite sign, whereby the latter acts as a safety net. Kumar and Seppi (1994) argue that informed traders use a combination of market and limit orders that is similar to uninformed investors in order to hide. Kaniel and Liu (2006) find that informed traders prefer limit orders over market orders when their informational advantage is long-lived. In the model by Goettler, Parlour, and Rajan (2009) traders first choose whether to acquire information, and then decide on their order. When the volatility of the fundamental value is low informed traders are shown to provide liquidity at the best quotes. According to Roşu (2015) informed traders prefer to submit limit orders when their informational advantage is relatively small.

[6]A major challenge is identifying which traders are informed in the first place. Beber and Caglio (2005) use the period before earnings announcements as a proxy for a period with increased information asymmetries, and assume that buyers were more likely to be informed because prices rose after announcements in their sample. They find that buyers more frequently submit non-aggressive limit orders on the bid side of the order book. Bloomfield, O'Hara, and Saar (2005) resort to an experimental setting to distinguish between informed traders and uninformed traders and show that informed investors use market orders in the beginning of the trading period, when their informational advantage is large, but switch to limit orders when the bulk of information has come to be incorporated in prices. Kaniel and Liu (2006) measure order informativeness as the conditional probability that the midquote increases (decreases) after a buy (sell) order, and show that for small and medium-sized orders limit orders contain more information than market orders.

bid price they may submit limit buy orders, when it is overvalued by the current ask price they can submit limit sell orders). The argument goes that this could cause imbalances in the state of the limit order book that are informative on the fundamental value of a security. However, this explanation also implies that informed traders do not succeed very well in hiding their informational advantage, and yet price impact is not instantaneous. This in itself is puzzling and requires prices to be inefficient for some reason.[7] One could also argue that informed traders would condition their order submission strategies on the current state of the limit order book in a way that allows them to hide within uninformed limit orders, which should make them less easily detectable than just by observing order book imbalances.

There is an alternative mechanism that allows the order book to contain fundamental information without necessarily assuming that a fraction of limit orders is submitted by informed traders. It only requires that rational uninformed traders adjust their liquidity provision on a discrete price grid depending on the information content assigned to the historical order flow. For instance, after a buy order they adjust their beliefs upward and revise their limit orders accordingly. They can do so either by submitting orders at higher bid and ask prices, or alternatively by reducing the depth offered at the best ask price relative to the best bid price. The latter is especially a valid strategy when the tick size is discrete (so prices cannot continuously adjust) and new information causes only a slight change in beliefs. Such behavior makes the depth imbalance contain fundamental information. However, for this to cause returns to be predictable it still implies that current bid and ask prices are not efficient.

Return predictability on the basis of limit order book information is in itself sometimes also considered as evidence that the order book contains fundamental information. However, such a conjecture does not take into account that the order book can also predict returns because it can generate temporary price deviations due to predictable order choice. Tombeur and Wuyts (2015) find evidence that this latter explanation for predictability is relatively more important, and that fundamental information in the limit order book contributes only to a lesser extent. Returns over longer time horizons are less predictable, indicating short-term price reversals. Furthermore, stocks that exhibit more informed trading have lower return predictability, which suggests that informed traders could actually help in removing short-term price pressure, rather than causing it.

[7] In models where rational risk-neutral traders are asymmetrically informed, information that is contained in the order flow is instantaneously reflected into market prices and does not contain any predictive power with regard to future prices, i.e. prices are efficient (e.g., Glosten and Milgrom, 1985). If the state of the limit order book contains fundamental information about future prices, then current prices are not efficient, and it takes somehow time for the market to process this information.

2.2.2 Return Predictability Across Venues

We now turn to the question of why order book information on one trading venue should predict returns on *another* trading venue. The simplest explanation is that prices for the same underlying cannot deviate too much because they are bound by an arbitrage relationship. Therefore, returns across venues should be correlated contemporaneously. A key role here may be for (HFT) market makers that are active on multiple venues for the same securities. Such traders contribute to the interlinking of different limit order books. Menkveld (2013) identifies a large HFT who employs essentially a market making strategy across two venues, the traditional listing exchange (Euronext) and an entrant venue (Chi-X). He finds that the entry of the HFT, who participates in more than 70 percent of trades on the entrant venue, coincides with a substantial increase in the liquidity and market share of this venue. This type of trader, who actively monitors multiple venues, causes information that originates on one venue to have an effect on prices of other venues. van Kervel (2015) shows (both theoretically and empirically) that after a trade is observed market makers revise their unexecuted limit orders on other venues in accordance with their updated beliefs. Also arbitrageurs can play a role in interlinking different venues. Foucault, Kozhan, and Tham (2015) show that fast traders exploit short-lived arbitrage opportunities when market makers' quotes are not updated simultaneously across markets.

Arbitrage across venues and a similar information impact thus provide an explanation of why returns across venues are contemporaneously correlated. But the question is whether returns should also be correlated when they represent temporary price pressure driven by order choice considerations. Because trading is anonymous it is impossible to distinguish between orders submitted (or canceled) by informed and uninformed traders. Therefore each order book event bears an information content. Market makers, or other traders interlinking different venues, can thus propagate both permanent and temporary price changes across these venues.

There is also a direct connection between the state of the order book on one venue and future returns on another venue. When the same asset is simultaneously traded on multiple venues, order choice also entails venue choice. As in the single venue setting, order choice depends on execution probability, and thus the state of *all* order books. In a static model without asymmetric information, Foucault and Menkveld (2008) show that the absence of time priority across limit order books allows traders to jump the queue (without increasing their order aggressiveness) by submitting limit orders in a competing limit order book. They also show that jumping the queue becomes relatively more attractive as more brokers employ smart order routing technology, which allows these brokers to trade in the competing limit order book. In their model the use of smart order routing technology increases the execution probability of orders in the entrant venue relative to that of the incumbent venue. The model does not directly allow to make empirical predictions concerning order book and price dynamics, but intuitively, queue-jumping between order books can explain return predictability across venues. For instance, suppose a patient buyer[8] encounters a relatively thick bid queue on venue j, while the bid price on venue i is one tick lower. Depending on the expected relative execution probabilities across venues (which depends on the fraction of brokers using smart order routers according to Foucault and Menkveld (2008)), it might be optimal to jump the queue at venue j by submitting a price-improving limit buy order at venue i (at the same bid price as venue j). This would result in a positive return on venue i. We therefore put forward the following hypothesis.[9]

Hypothesis 1 *Returns on venue i can be predicted by the lagged state of the order book of venue $j \neq i$. In particular, a positive (negative) imbalance between liquidity at the bid and ask side on venue j predicts positive (negative) returns on venue i.*

Panel (a) of Figure 2.1 illustrates the previous example graphically. The figure shows, for two given hypothetical order books on venue i and venue j, the 9 potential situations for their *relative position* vis-à-vis each other. The solid lines represent the best depth queues on venue i (which are balanced), while the dotted lines represent best depth queues on venue j (which show a positive order book imbalance). As can be judged from the figure, the relationship between order book imbalances and returns across venues is likely to depend on the relative position of the bid and ask quotes on the price grid. The relative position of quotes on the same side of the market (the bid side for buyers and the ask side for sellers) determines whether incoming traders can jump the queue on venue j by submitting orders on venue i. This can be a useful strategy for patient buyers in the situation depicted in Panel (a), as venue j is shown to be very competitive in comparison to venue i. But it is less likely for the situation from Panel (i), in which the reverse is true. Here, jumping the queue on venue j does not cause a positive midquote return on venue i. Instead, the opposite (traders jumping the queue at venue i) is more plausible. Panel (e) depicts an intermediate scenario in which queues of both venues are at the same prices. In sum, what matters for traders when they decide to queue-jump between venues is not only the size of the queue, but also their relative position in the price grid. Ceteris paribus, queue-jumping should be more effective when traders can submit orders on a venue that is quoting less competitive prices.

[8] A patient trader here refers to a trader who is more likely to submit a limit order because his personal valuation of the asset is not 'extreme', and therefore his gains from trade are not high enough to submit a market order. Patient traders therefore supply liquidity, i.e., they are market makers.

[9] For clarity of exposition, in the remainder of the text we refer to the venue on which we want to predict returns as **venue i**, and another competing venue of which we use limit order book information as **venue j**.

Figure 2.1: Order Book Imbalances and Quote Positions Across Venues

This figure shows the nine potential scenarios for the relative position of bid and ask prices on venue j relative to venue i. The vertical axis displays the bid and ask prices, while the horizontal axis shows the number of shares offered at these prices. The solid lines represents the depth (in the number of shares) supplied at bid and ask prices on venue i (in this example always €9.98 and €10.02, respectively). The dotted lines represent the depth supplied at bid and ask prices of venue j. The depth does not vary between scenarios. Venue i always has a zero order book imbalance, while venue j has a positive order book imbalance. The position of bid and ask prices on the price grid of venue j changes between potential scenarios. The first row always presents a situation in which the best ask price of venue j is better than that of venue i; in the second row the ask prices are equal; and in the third row the ask price of venue j is lower than that of venue i. The first column always presents a situation in which the best bid price of venue j is better than that of venue i; in the second column bid prices are equal; and in the third column the bid price of venue j is always worse than that of venue i.

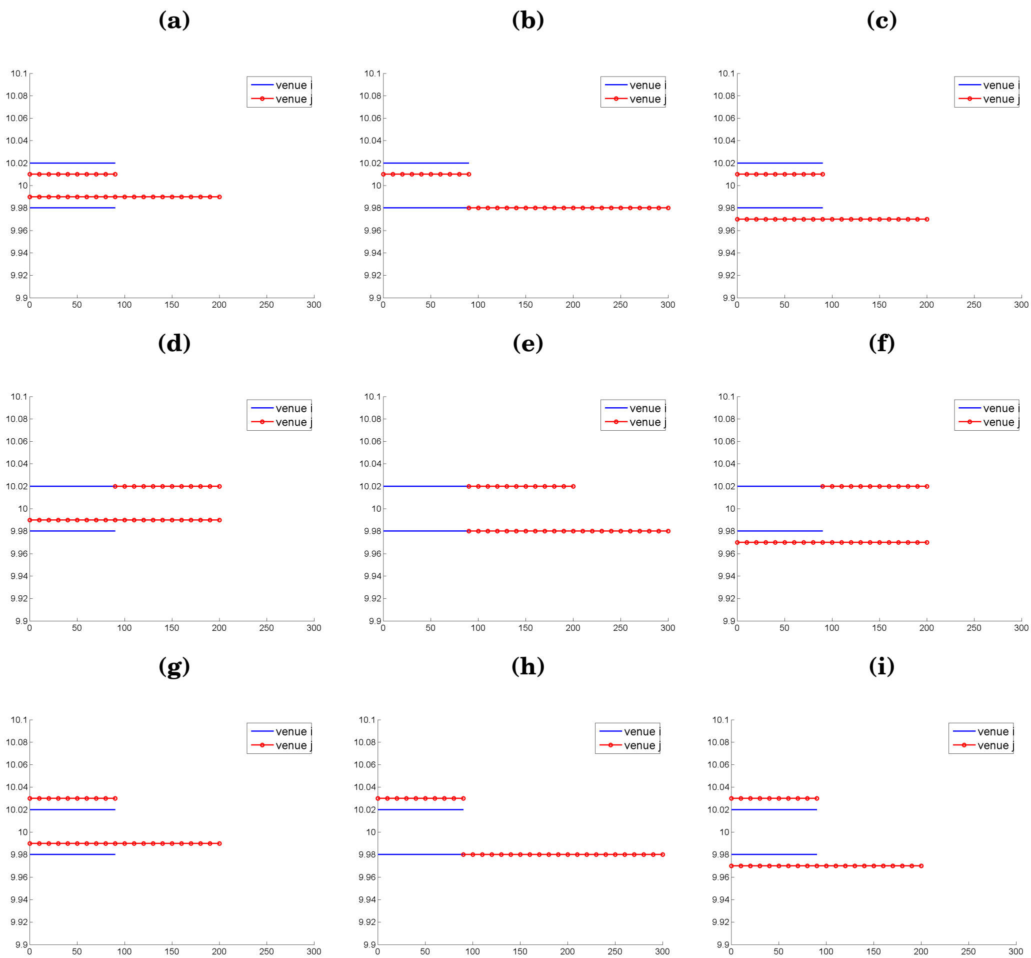


Figure 2.1 also shows that taking into account the relative position of the quotes on both venues is complicated because the position need not be symmetric at the bid and ask side. Both in Panels (a) and (g), for instance, we expect that patient buyers on venue j would want to skip the bid queue by submitting a price improving limit order on venue i. For patient sellers on venue j, in Panel (g), all else equal, they would prefer to submit a price-improving sell limit order on venue j itself rather than adding to the queue of venue i. So for the situation in Panel (g) we would unambiguously predict positive returns on venue i. By contrast, in Panel (a), although the depth queue is relatively small, patient sellers on venue j can be tempted to jump the queue and submit a price-improving sell limit order on venue i, which would cause a decrease in the (mid)quote of venue i. Our description here is intuitive, but it captures the notion that the relationship between the state of the order book on one venue and future returns on another venue does not depend only on the relative depth of the bid and ask side, but also on the relative position of the venue quotes, or the *competitiveness* of their prices. Moreover, the relative position of the quotes itself contains information that can be used to predict returns. This needs to be taken into account for the empirical setup. Finally, note that even without competition or queue-jumping between limit order books prices would tend to adjust to each other across venues since they are for the same asset that is traded on all these venues, and thus prices are bound by a no-arbitrage condition. Our conjecture that the position of the quotes on the price grid matters for return predictability leads to the following two hypotheses.

Hypothesis 2 *The positive relationship between returns on venue i and the lagged state of the limit order book on venue $j \neq i$ is the strongest (weakest) when the best quote on venue $j \neq i$ is strictly better (worse) than the best quote of venue i.*

Hypothesis 3 *There exists a positive relationship between returns on venue i and the lagged distance between the midquotes of venue i and venue $j \neq i$, where distance is measured as the (signed) difference in midquotes. A symmetrical effect exists for venue j.*

2.2.3 The Effect of Fragmentation

Foucault and Menkveld (2008) argue that there is an important role for smart order routing technology employed by brokers to connect different venues and make competition among them viable. In their setting, when no brokers employ smart order routing technology there is no incentive for liquidity providers to jump the queue on the incumbent venue by submitting a limit order and provide liquidity on the entrant. As more brokers use smart order routers the execution probability of limit orders on the entrant venue increases, and so does the incentive to jump the queue. As such, the trader avoids having to join a queue on the incumbent venue and becomes first in the queue on the entrant. This suggests that the expected execution probability of limit orders matters when traders decide whether to jump the queue. As a rough proxy of expected execution probability on a venue we use its market share of trading volume (with hindsight). When a trading venue j become more relevant for traders, the order book of venue j becomes a more important determinant for order choice, and thus its contribution for return predictability becomes larger.

Hypothesis 4 *The positive relationship between returns on venue i and the lagged state of the limit order book on venue j, is stronger when venue j has a larger market share of trading volume.*

2.3 Methodology

2.3.1 Return Predictability on the Own Venue

We first establish that returns are predictable on each venue i using only lagged order book imbalances of the own venue b^i_{t-1} as a predictor.

$$r^i_t = \alpha + \beta^i b^i_{t-1} + u^i_t \tag{2.1}$$

Log returns r^i_t are calculated from midquote prices m^i_t. Because our hypotheses from Section 2.2 imply a symmetric prediction for both the best bid and ask price we use the midquote as a shortcut. If returns are predictable because the order book contains information the best bid and ask price are impacted the same. Return predictability due to order choice considerations similarly implies symmetric effects on potential buyers and sellers. For instance, if buyers are predicted to become more aggressive, then sellers should become less aggressive. In addition, the midquote price is often used as a benchmark price by investors for measuring transaction costs, as it is considered to be an estimate of the fundamental value of a security.

Order book imbalances b^i_t measure the difference between depth (in number of shares) quoted on the bid side $D^{i,bid}_t$ and on the ask side $D^{i,ask}_t$. Following previous research we scale the imbalance by total depth offered on both sides of the market.

$$b^i_t = \frac{D^{i,bid}_t - D^{i,ask}_t}{D^{i,bid}_t + D^{i,ask}_t} \tag{2.2}$$

We use depth at the best prices because Cao, Hansch, and Wang (2009) and Tombeur and Wuyts (2015) show that liquidity at the top of the book is the best predictor of future returns.

We measure t in calendar time and use a 10 second sampling frequency. Traditionally longer time intervals have been used to predict returns (5 minutes is common), but the advent of high-frequency trading has drastically increased order book quoting activity. Because order book dynamics are the drivers of return predictability, we assume that if returns are predictable on the basis of order book information, the relation is more outspoken at shorter time horizons.[10] The model is estimated per stock and per day.

Venues i for which we predict returns include both the listing exchange, Euronext (Enx), and alternative venues, Chi-X (Chi), Turquoise (Tur) and Bats (Bat). In addition we also investigate return predictability at the level of the consolidated market (Con). To do so, we calculate a European Best Bid and Offer (EBBO), i.e. the best bid and ask prices across all venues, at each time t, and from these prices a consolidated midquote price m_t^{Con}. We also calculate the order book imbalance at the consolidated level b_t^{Con}.[11]

2.3.2 Return Predictability Across Venues

We extend the baseline model of Equation (2.1) to include order book imbalances from other venues.

$$r_t^i = \alpha + \beta^i b_{t-1}^i + \beta^{Con\backslash i} b_{t-1}^{Con\backslash i} + u_{i,t} \tag{2.3}$$

where i represents the venue for which we predict returns, and $Con\backslash i$ is the consolidated market excluding venue i. This extended model allows us to see whether order book imbalances from other venues can predict returns on venue i, after controlling for the order book imbalance of venue i itself. This model is therefore considered to be a test of hypothesis 1.

Furthermore, we try to disentangle which venues contribute the most to predictability of returns by estimating the following model for each venue i.

$$r_t^i = \alpha + \beta^i b_{t-1}^i + \sum_{j \neq i} \beta^j b_{t-1}^j + u_{i,t} \tag{2.4}$$

Hypothesis 2 states that the predictive power of the order book state for future returns depends not only on the imbalance between the bid and ask side, but also on the relative position of the quotes of the different venues on the price grid. When venue j is quoting a higher bid (lower ask) than venue i, patient traders have relatively more incentives to jump the queue on venue j and increase their

[10] The choice of the sampling frequency entails a trade-off. When we choose intervals that are too long, the relation between the lagged state of the order book and returns becomes weaker. Multiple traders may enter the market and trade for liquidity or informational reasons, and as such the temporary price effect of the order book state disappears. When we choose intervals that are too short, there could be multiple periods in which no traders enter the market, and the relation between the lagged state of the order book also obscures. Ultimately, the optimal interval length may depend on the frequency with which traders enter the market. Across all the stocks in our sample we believe 10 seconds balances the desire to have traders entering the market with a reasonable probability in any time interval, without having *too much* activity.

[11] At the level of the individual venues we remove observations for which there is only a price quoted on one side of the order book (because we cannot calculate a meaningful midquote or imbalance), which is a situation that can occur sometimes for less actively traded stocks on MTFs. At the consolidated level we allow for the possibility that these one-sided quotes are (part of) the EBBO.

liquidity provision on venue i. Hence, more competitive quotes cause the limit order book of venue j to have a stronger predictive power for venue i returns. We test this hypothesis by decomposing depth from competing venues into depth that is quoted at better prices, depth quoted at the same price and depth quoted at worse prices, relative to venue i. In particular, at each point in time t, for each venue i we classify bid depth from all competing venues $j \neq i$ (which constitute the consolidated market excluding i, $Con\backslash i$) as follows:

$$\begin{aligned}
D_t^{Con\backslash i(B),bid} &= \sum_{j\neq i} D_t^{j,bid}\mathbb{1}(P_t^{j,bid} > P_t^{i,bid}) \\
D_t^{Con\backslash i(S),bid} &= \sum_{j\neq i} D_t^{j,bid}\mathbb{1}(P_t^{j,bid} = P_t^{i,bid}) \\
D_t^{Con\backslash i(W),bid} &= \sum_{j\neq i} D_t^{j,bid}\mathbb{1}(P_t^{j,bid} < P_t^{i,bid})
\end{aligned}$$

Ask depth is classified similarly. Superscripts $Con\backslash i(B)$, $Con\backslash i(S)$ and $Con\backslash i(W)$ refer to depth at better prices, the same prices and worse prices than venue i, respectively. Using depth from competing venues classified accordingly we construct the following three order book imbalance measures.

$$\begin{aligned}
b_t^{Con\backslash i(B)} &\begin{cases} =0 & if \quad D_t^{Con\backslash i(B),bid} + D_t^{Con\backslash i(B),ask} = 0 \\ =\dfrac{D_t^{Con\backslash i(B),bid} - D_t^{Con\backslash i(B),ask}}{D_t^{Con\backslash i(B),bid} + D_t^{Con\backslash i(B),ask}} & otherwise \end{cases} \\
b_t^{Con\backslash i(S)} &\begin{cases} =0 & if \quad D_t^{Con\backslash i(S),bid} + D_t^{Con\backslash i(S),ask} = 0 \\ =\dfrac{D_t^{Con\backslash i(S),bid} - D_t^{Con\backslash i(S),ask}}{D_t^{Con\backslash i(S),bid} + D_t^{Con\backslash i(S),ask}} & otherwise \end{cases} \\
b_t^{Con\backslash i(W)} &\begin{cases} =0 & if \quad D_t^{Con\backslash i(W),bid} + D_t^{Con\backslash i(W),ask} = 0 \\ =\dfrac{D_t^{Con\backslash i(W),bid} - D_t^{Con\backslash i(W),ask}}{D_t^{Con\backslash i(W),bid} + D_t^{Con\backslash i(W),ask}} & otherwise \end{cases}
\end{aligned} \tag{2.5}$$

To test Hypothesis 2 we then estimate the following model.

$$\begin{aligned}
r_t^i &= \alpha + \beta^i b_{t-1}^i + \beta^{Con\backslash i(B)} b_{t-1}^{Con\backslash i(B)} + \beta^{Con\backslash i(S)} b_{t-1}^{Con\backslash i(S)} \\
&\quad + \beta^{Con\backslash i(W)} b_{t-1}^{Con\backslash i(W)} + u_{i,t}
\end{aligned} \tag{2.6}$$

According to this hypothesis, $\beta^{j(B)}$ should be more positive than $\beta^{j(S)}$, while $\beta^{j(W)}$ is implied to be the weakest effect.

To test for a direct effect of the relative positive of the quotes on the price grid (in addition to the effect of order book imbalances), as hypothesized in Hypothesis 3, we introduce a measure of relative *distance* between midquotes. We measure, from the point of view of venue i, the distance from its midquote m_t^i to the midquote m_t^j of venue j as follows

$$d_t^j = \frac{m_t^j - m_t^i}{(\frac{1}{2})(m_t^j + m_t^i)} \tag{2.7}$$

The distance is expressed relative to the average midquote of both venues. As before we construct for each venue i the consolidated order book of the competing venues ($Con\backslash i$). We then extend model (2.3) by adding the distance between the midquote from the best alternative venue j and venue i. Hypothesis 3 predicts a positive δ^j, as prices on different venues tend to adjust to each other.

$$r_t^i = \alpha + \beta^i b_{t-1}^i + \beta^{Con\backslash i} b_{t-1}^{Con\backslash i} + \delta^{Con\backslash i} d_{t-1}^{Con\backslash i} + u_t^i \tag{2.8}$$

And then we disentangle, similarly as before, which venues contribute to predictability the most.

$$r_t^i = \alpha + \beta^i b_{t-1}^i + \sum_{j \neq i} \beta^j b_{t-1}^j + \sum_{j \neq i} \delta^j d_{t-1}^j + u_t^i \tag{2.9}$$

2.3.3 The Effect of Fragmentation

The predictive ability of the state of the order book of a venue for returns on another venue is likely to be different across venues, across stocks and over time. Hypothesis 4 states that part of the differences across trading venues and over time could be explained by fragmentation of trading volume across venues. In particular, an order book of a trading venue is expected to be better predictor of returns on another trading venue when it has a larger market share. A larger market share for a venue indicates that it is more relevant, and thus more connected to the consolidated market, as traders route more orders to these venues. Under such circumstances, a venue is more likely to matter. To test this hypotheses we estimate the following models for each venue i and competing venue j in the sample separately.

$$\begin{aligned} r_t^i &= \alpha + \beta^i b_{t-1}^i + \beta^j b_{t-1}^j + \delta^j d_{t-1}^j + \mu^j ms_{t-1,\tau}^j \\ &\quad + \gamma^j b_{t-1}^j * ms_{t-1,\tau}^j + \lambda^j d_{t-1}^j * ms_{t-1,\tau}^j + u_t^i \end{aligned} \tag{2.10}$$

where $ms_{t-1,\tau}^j$ is the relative market share of venue j, considering only venue j and venue i, on day τ. Important to note here is that this model is not a predictive model, as we use hindsight to determine market shares. Because we aggregate market shares to the daily level we can no longer estimate the model per day. Instead, we estimate the model using the full one year time series. The implicit assumption is that all variability in predictive ability (and thus coefficients) across days of both the order book imbalance b^j and the distance d^j can be captured by the relative market share of the venues ms^j_τ.

2.4 Data

2.4.1 Institutional Background

Today's market structure has evolved from a largely consolidated system into a market where order flow is fragmented over different trading venues. This holds for U.S. markets, as well as European markets. MiFID is the legal framework in which financial markets operate inside the European Economic Area since November 2007. It has expanded the set of regulated trading venues from the Regulated Markets (RMs), which encompass the traditional national stock exchanges, to also include Multilateral Trading Facilities (MTFs) and Systematic Internalizers (SIs). Both RMs and MTFs are multilateral trading venues that match third-party order flow. Most of these venues are subject to pre-trade transparency requirements (with some room for exceptions[12]) and thus operate public limit order books that display a large amount of the orders that are submitted to their books. MTFs are similar to Electronic Communications Networks (ECNs) in the U.S.

Our dataset contains Belgian, Dutch and French index stocks. Although they have different nationalities, they have in common that they are all listed on one of the Euronext exchanges (Brussels, Amsterdam or Paris), and are mainly traded on the same four trading venues during our sample period: Euronext, Chi-X, Turquoise, Bats and Nasdaq OMX.[13] Chi-X started its operations in April 2007 and has been the most successful MTF in Europe. Turquoise, Bats and Nasdaq OMX were all launched during the second half of 2008. These venues can differ in a number of ways (e.g., fee structure, clearing house, latency or order types), but in general they are quite similar; all operate on the basis of a public limit order book.

2.4.2 Data and Sample

From the full sample of 85 French, Dutch and Belgian index stocks we select a subset of 30 stocks, 10 of each listing exchange (Paris, Amsterdam, Brussels). On each exchange we sort stocks into quintiles based on their level of fragmentation

[12]Four types of pre-trade transparency waivers can be granted which have allowed opaque trading to flourish in Europe in recent years. In addition, under certain conditions, MiFID allows opaque transactions on the unregulated Over-the-Counter (OTC) market. Furthermore, post-trade transparency obligations can be delayed for orders that are large in scale as compared to normal market size.

[13]Nasdaq OMX Europe seized trading in May 2010 because it did not attract a large enough market share. It is therefore not included in the sample. Stocks can also be cross-listed on other venues, such as the London Stock Exchange, New York Stock Exchange, Luxembourg Stock Exchange or Bolsa de Madrid. Because the shares, currency and trading hours are generally different we do not include trading on these venues in our analysis.

of trading volume, and from each of those quintiles we pick the stocks with the lowest and highest amount of trading volume. This approach ensures there is variation on the dimensions of the listing exchange, market fragmentation and overall trading activity. We use a one year time series for each stock, from October 2009 until September 2010.

Data are obtained from Thomson Reuters Tick History. The dataset has two parts. The first part of the dataset contains trade and quote data from all venues, timestamped to the millisecond. The second part contains order book data for these venues, also timestamped to the millisecond, identifying prices and visible depth up to the tenth best price available at every update in the limit order book. A new update is disseminated each time a trader submits, cancels or modifies an order at one of these prices. We apply standard data cleaning procedures (e.g., remove updates outside opening hours of the market, delete negative spreads, ...) and use the updates to construct snapshots of the limit order books at pre-specified time intervals (10 seconds) on each venue. For comparability reasons across days we also remove trading days during which a stock is not traded for a full day during exchange opening hours (i.e. December 24 and 31 for all stocks, and days on which there is no activity for more than 20 minutes at the best prices on Euronext, as these are likely trading halts).

2.4.3 Descriptive Statistics

Table 2.1 presents for each of the 30 stocks in our sample the time series mean for a selection of variables. Panel A presents the means for variables at the consolidated market level for each stock. In the bottom six lines of the table we show the cross-sectional average and median values for the full stock sample, as well as for the five stocks experiencing the highest fragmentation (HFrag sample) and lowest fragmentation (LFrag sample). Fragmentation of trading volume across the four lit venues in our sample is measured as one minus the Herfindahl-Hirschman Index (Degryse, de Jong, and van Kervel, 2015). The average (median) stock has a fragmentation of 0.45 (0.48). High fragmentation stocks have an average (median) fragmentation of 0.53 (0.52), while stock trading is much more concentrated for low fragmentation stocks, with an average (median) fragmentation of 0.29 (0.31). The average (median) stock in the sample has a market capitalization of 20 billion (6 billion) euro. The stocks listed on Euronext Paris are in general larger in size, while those listed in Brussels are much smaller. Similar observations can be made for trading volume, liquidity (spread and depth), quoting activity (the number of order book updates at the best prices, which includes order submissions, cancellations and trades) and the average time since the last quote update for each snapshot of the limit order book. The latter is larger than the average duration between quote updates (as implied by the number of quote updates) because there are both active and calm periods in the order book, and quote updates also tend to cluster in time. The

Table 2.1: Descriptive Statistics

This table presents summary statistics (time series averages) for each stock in the sample based on daily observations for a selection of market variables. The last six lines of the table show the cross-sectional averages and medians for all stocks (All), the five stocks that have the highest fragmentation (HFrag), and the five stocks that have the lowest fragmentation (LFrag). Market capitalization is expressed in million euro, volume in thousand euro. Price is the average midquote price (in euro). Spread is expressed in basis points, relative to the prevailing midquote. Abs Ret is the average absolute 10-second midquote return (expressed in basis points). Depth is the sum of bid and ask depth at the best prices in the consolidated market, expressed in euro. Daily observations of the price, spread and depth are time-weighted averages based on 10 second snapshots throughout the trading day. The number of quotes (NQuotes) is the number of times the either the best bid and ask prices, or depth, at the consolidated market changes within the trading day. The time difference since the last quote update in the consolidated market (TimeLast) is the average time difference between a snapshot of the consolidated limit order book and the last time it was updated because an order was submitted, modified or canceled at the best prices (measured in seconds). Fragmentation (Frag) of trading volume is measured as one minus the Herfindahl-Hirschman Index (HHI) based on trading volume market shares of the four venues in the sample.

		Panel A (1)			
Ticker	Stock	Listing	Frag	Market Cap	Volume
BNPP	BNP Paribas	PA	0.47	53,982	316,046
EDF	Electricité de France	PA	0.48	11,164	70,041
LAGA	Lagardère Group	PA	0.47	3,401	21,183
LVMH	LVMH	PA	0.48	21,969	125,455
SASY	Sanofi	PA	0.52	56,168	230,262
SEVI	Suez Environnement	PA	0.50	3,892	18,669
SOGN	Société Générale	PA	0.46	28,470	254,743
STM	STMicroelectronics	PA	0.47	3,686	38,540
TCFP	Thales	PA	0.43	2,805	17,964
TOTF	Total	PA	0.52	96,112	358,240
COR	Corio	AS	0.49	3,770	16,380
ELSN	Reed Elsevier	AS	0.48	6,343	33,941
FUGR	Fugro	AS	0.40	3,441	23,086
HEIN	Heineken	AS	0.51	17,944	62,186
ING	ING	AS	0.42	25,937	264,734
ISPA	ArcelorMittal	AS	0.50	42,420	367,768
PHG	Philips	AS	0.52	21,867	151,772
RDSa	Royal Dutch Shell	AS	0.50	75,211	203,427
SBMO	SBMO	AS	0.43	2,224	17,987
UN	Unilever	AS	0.56	34,178	176,198
ABI	AB Inbev	BR	0.50	60,319	111,834
ACKB	Ackermans & Van Haaren	BR	0.31	1,741	1,905
BEFB	Befimmo	BR	0.30	1,006	1,423
BEKB	Bekaert	BR	0.31	2,544	7,806
FOR	Fortis/Ageas	BR	0.34	5,776	28,886
KBC	KBC	BR	0.40	11,909	33,835
MSTAR	Mobistar	BR	0.42	2,704	9,803
NYR	Nyrstar	BR	0.22	923	8,593
OMEP	Omega Pharma	BR	0.32	841	2,036
UCB	UCB	BR	0.45	5,256	10,852
Average All			0.44	20266.74	99519.87
Median All			0.47	6059.55	33887.90
Average HF			0.53	45,254	195,732
Median HF			0.52	34,178	176,198
Average LF			0.29	1,411	4,353
Median LF			0.31	1,006	2,036

Table 2.1 continued.

Panel A (2)

Ticker	Price	Spread	Depth	NQuotes	TimeLast
BNPP	52.50	3.47	125,595	204,807	1.11
EDF	37.04	3.84	65,358	82,799	2.43
LAGA	28.90	8.01	52,167	60,872	4.14
LVMH	84.43	4.43	96,101	121,594	1.96
SASY	51.27	3.17	161,096	132,176	1.69
SEVI	15.24	7.12	71,122	43,857	5.32
SOGN	43.22	4.05	80,303	189,960	1.23
STM	6.33	6.02	47,018	87,502	2.95
TCFP	30.00	8.03	42,105	28,005	9.34
TOTF	40.76	2.50	134,595	176,198	1.09
COR	45.15	9.70	42,287	42,485	5.91
ELSN	8.77	4.74	46,561	53,675	4.17
FUGR	43.39	8.58	43,689	47,794	5.21
HEIN	34.75	4.48	90,954	54,993	3.95
ING	7.57	4.40	90,970	204,794	1.20
ISPA	27.11	3.62	147,566	200,195	1.05
PHG	22.37	4.67	156,935	100,724	2.31
RDSa	21.21	3.16	444,890	161,203	1.50
SBMO	13.57	9.57	72,947	38,208	4.89
UN	22.02	3.40	219,110	139,882	1.72
ABI	37.61	5.31	93,599	85,877	2.86
ACKB	51.98	24.05	25,205	28,192	32.10
BEFB	59.92	25.14	25,973	14,804	45.53
BEKB	128.32	15.29	47,677	21,803	15.85
FOR	2.46	11.91	57,343	67,813	7.51
KBC	33.36	13.29	47,384	41,030	20.28
MSTAR	45.01	9.69	34,680	24,472	11.46
NYR	9.24	21.62	57,388	16,575	15.52
OMEP	34.67	18.58	25,750	25,184	36.60
UCB	28.64	10.71	42,175	24,323	11.40
Average All	35.56	8.75	89618.22	84059.83	8.74
Median All	34.01	6.57	61373.00	57932.53	4.16
Average HF	34.23	3.64	152,538	120,795	2.15
Median HF	34.75	3.40	156,935	132,176	1.72
Average LF	56.83	20.93	36,399	21,311	29.12
Median LF	51.98	21.62	25,973	21,803	32.10

Table 2.1 continued.

Note: Panel B shows statistics for the four different trading venues in the sample: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). Market Share is the market share of daily trading volume. NQuotes (Relative) shows the number of times the best prices or depths of the limit order book at each venue is updates, relative to the total amount of best order book updates at all markets. TimeLast shows the average time difference between a snapshot of the limit order book and the last time it was updated because an order was submitted, modified or canceled at the best prices (measured in seconds). Distance (Ask) is the distance between the best ask price of the order book, and the best ask price in the consolidated market (measured in basis points relative to the ask price). Depth (Relative) shows the number of shares available at the best prices of each venue, relative to the total amount of shares available at the best prices of all venues.

Panel B (1)

Ticker	Market Share				NQuotes (Relative)				TimeLast			
	Enx	Chi	Tur	Bat	Enx	Chi	Tur	Bat	Enx	Chi	Tur	Bat
BNPP	0.69	0.23	0.03	0.05	0.23	0.32	0.22	0.23	14.33	2.28	6.41	3.65
EDF	0.67	0.25	0.04	0.03	0.25	0.35	0.19	0.22	16.45	5.56	13.34	9.63
LAGA	0.69	0.21	0.06	0.03	0.27	0.28	0.23	0.22	20.55	11.17	16.87	15.68
LVMH	0.67	0.24	0.05	0.04	0.24	0.33	0.25	0.18	16.23	4.42	8.53	8.12
SASY	0.64	0.25	0.05	0.06	0.25	0.31	0.20	0.23	14.18	3.73	8.81	5.85
SEVI	0.65	0.24	0.07	0.04	0.30	0.30	0.19	0.21	20.97	11.74	21.78	20.19
SOGN	0.70	0.22	0.04	0.04	0.21	0.32	0.26	0.21	13.59	2.84	5.59	4.29
STM	0.70	0.18	0.07	0.05	0.19	0.31	0.22	0.27	18.95	7.57	12.14	9.05
TCFP	0.72	0.20	0.05	0.03	0.29	0.28	0.23	0.21	28.97	21.98	31.69	30.13
TOTF	0.64	0.25	0.04	0.07	0.25	0.29	0.20	0.27	14.13	2.71	6.05	3.44
COR	0.66	0.23	0.08	0.03	0.26	0.25	0.31	0.18	23.77	17.94	24.15	24.76
ELSN	0.68	0.23	0.06	0.04	0.27	0.27	0.23	0.22	19.80	9.97	20.02	15.07
FUGR	0.75	0.16	0.05	0.03	0.28	0.25	0.27	0.20	20.76	13.88	22.33	18.83
HEIN	0.65	0.24	0.06	0.05	0.28	0.30	0.23	0.19	18.59	9.55	19.33	20.44
ING	0.73	0.20	0.03	0.04	0.22	0.28	0.25	0.26	13.51	3.06	9.90	4.42
ISPA	0.66	0.24	0.03	0.07	0.19	0.32	0.25	0.24	14.02	2.52	9.82	3.37
PHG	0.64	0.26	0.04	0.06	0.27	0.34	0.18	0.20	15.31	5.51	14.53	8.63
RDSa	0.66	0.25	0.03	0.05	0.28	0.28	0.15	0.29	14.02	3.80	13.05	4.78
SBMO	0.73	0.19	0.04	0.04	0.34	0.24	0.21	0.22	20.01	14.23	23.86	26.82
UN	0.60	0.27	0.04	0.09	0.25	0.28	0.14	0.33	14.67	4.71	13.82	5.47
ABI	0.64	0.27	0.05	0.04	0.20	0.34	0.22	0.24	17.50	6.20	18.46	10.52
ACKB	0.81	0.12	0.04	0.03	0.24	0.31	0.25	0.20	75.37	75.27	155.03	131.39
BEFB	0.82	0.11	0.05	0.02	0.30	0.30	0.21	0.18	93.46	116.03	228.67	225.69
BEKB	0.81	0.13	0.03	0.03	0.29	0.29	0.24	0.18	40.67	39.31	84.46	86.84
FOR	0.80	0.14	0.03	0.04	0.15	0.34	0.15	0.35	29.84	17.86	45.02	28.62
KBC	0.74	0.19	0.04	0.03	0.28	0.31	0.26	0.16	43.95	28.12	46.96	80.76
MSTAR	0.72	0.21	0.04	0.03	0.28	0.33	0.26	0.13	34.27	25.41	42.55	70.03
NYR	0.88	0.08	0.03	0.01	0.30	0.27	0.22	0.21	38.82	42.02	65.27	75.61
OMEP	0.80	0.12	0.05	0.03	0.24	0.28	0.30	0.17	83.80	83.17	97.63	201.28
UCB	0.70	0.23	0.04	0.04	0.28	0.35	0.21	0.16	33.98	23.60	65.54	62.81
Average All	0.71	0.20	0.05	0.04	0.26	0.30	0.22	0.22	28.15	20.54	38.39	40.54
Median All	0.69	0.22	0.04	0.04	0.26	0.30	0.22	0.21	19.91	10.57	19.67	17.25
Average HF	0.63	0.25	0.05	0.06	0.26	0.30	0.19	0.24	15.38	5.24	12.51	8.77
Median HF	0.64	0.25	0.04	0.06	0.25	0.30	0.20	0.23	14.67	4.71	13.82	5.85
Average LF	0.83	0.11	0.04	0.02	0.28	0.29	0.25	0.19	66.43	71.16	126.21	144.16
Median LF	0.81	0.12	0.04	0.03	0.29	0.29	0.24	0.18	75.37	75.27	97.63	131.39

Table 2.1 continued.

Panel B (2)

Ticker	Distance (Ask)				Depth (Relative)			
	Enx	Chi	Tur	Bat	Enx	Chi	Tur	Bat
BNPP	0.53	0.60	8.66	2.89	0.41	0.27	0.16	0.17
EDF	0.75	0.50	3.36	6.12	0.47	0.23	0.13	0.17
LAGA	0.58	1.01	3.57	9.24	0.43	0.21	0.20	0.17
LVMH	0.52	0.61	3.63	2.79	0.44	0.23	0.17	0.16
SASY	0.35	0.53	2.42	1.88	0.44	0.26	0.15	0.15
SEVI	0.61	1.08	3.83	3.67	0.45	0.26	0.16	0.14
SOGN	0.70	0.70	2.73	4.99	0.44	0.24	0.15	0.17
STM	0.99	1.31	3.01	3.12	0.39	0.19	0.23	0.19
TCFP	1.01	1.19	4.10	4.29	0.42	0.22	0.20	0.15
TOTF	0.39	0.48	2.32	1.23	0.42	0.27	0.16	0.16
COR	0.91	0.93	1.67	8.43	0.40	0.21	0.20	0.18
ELSN	0.65	0.82	1.91	2.78	0.46	0.23	0.17	0.14
FUGR	0.82	1.05	2.15	2.97	0.45	0.21	0.17	0.17
HEIN	0.44	0.57	2.02	5.11	0.45	0.25	0.16	0.14
ING	0.70	0.78	3.35	2.63	0.43	0.26	0.15	0.15
ISPA	0.54	0.64	4.12	1.71	0.36	0.25	0.20	0.19
PHG	0.40	0.59	2.60	1.98	0.45	0.28	0.13	0.13
RDSa	0.26	0.36	1.80	0.91	0.36	0.30	0.16	0.17
SBMO	0.76	1.79	5.81	9.09	0.45	0.22	0.17	0.16
UN	0.35	0.57	2.04	1.05	0.43	0.28	0.11	0.19
ABI	0.71	0.81	6.27	4.93	0.39	0.26	0.17	0.18
ACKB	1.25	8.73	16.72	29.94	0.27	0.20	0.20	0.32
BEFB	1.61	12.07	13.82	17.10	0.24	0.22	0.23	0.26
BEKB	0.97	5.91	26.37	24.98	0.29	0.20	0.24	0.25
FOR	1.36	3.63	26.82	12.06	0.40	0.22	0.18	0.19
KBC	1.49	4.07	6.34	24.98	0.37	0.22	0.16	0.24
MSTAR	0.74	1.63	7.96	18.23	0.35	0.19	0.19	0.26
NYR	0.93	10.21	14.40	29.63	0.45	0.19	0.15	0.21
OMEP	1.13	10.48	17.00	50.04	0.30	0.19	0.20	0.27
UCB	0.80	1.32	15.03	14.32	0.34	0.21	0.21	0.23
Average All	0.77	2.50	7.19	10.10	0.40	0.23	0.18	0.19
Median All	0.72	0.97	3.73	4.96	0.42	0.23	0.17	0.17
Average HF	0.39	0.55	2.28	2.25	0.44	0.27	0.14	0.15
Median HF	0.39	0.57	2.32	1.88	0.44	0.27	0.15	0.15
Average LF	1.18	9.48	17.66	30.34	0.31	0.20	0.21	0.26
Median LF	1.13	10.21	16.72	29.63	0.29	0.20	0.20	0.26

HFrag stocks are much larger in size, have higher trading volumes, are more liquid, have more frequent quote updates and a smaller average time since the last quote update than the LFrag stocks. The HFrag stocks are listed in Paris (SASY, TOTF) or Amsterdam (HEIN, PHG and UN), while the LFrag stocks are all listed in Brussels (ACKB, BEFB, BEKB, NYR and OMEP). In interpreting the results we point to differences between the HFrag and LFrag stocks, but bear in mind that these two subsamples differ on a wide array of characteristics.

Panel B of Table 2.1 presents means of variables that are measured on the different venues: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). Euronext has by far the largest market share (74% on average), while Chi-X is the largest competitor with a 20% market share for the average stock. Turquoise and Bats are relatively small, with a 4% average market share each. The dominance of Euronext is naturally more outspoken for LFrag stocks (83% market share) than for HFrag stocks (63% market share). For HFrag stocks Chi-X has a 25% average market share, while it is only 11% for LFrag stocks. For Turquoise and Bats the difference in market shares between HFrag stocks (5% and 6%) and LFrag stocks (4% and 2%) is smaller.

Differences in quoting activity are less outspoken across venues, but the pattern is rather similar. Chi-X appears to be the most active quoting venue, with 30% of the total amount of quotes for the average and median stock, followed by Euronext (26%), and finally Turquoise and Bats, with both 22% of quoting activity. The fact that quoting activity is more similar across venues, is in line with our expectations. When market makers adjust their liquidity provision on one venue, e.g., in response to an information event, they should also adjust their orders on other venues to avoid these becoming stale.

Liquidity is distributed similarly across venues. Both Euronext and Chi-X quote prices that are generally close to the consolidated best price, at 0.77 and 2.50 basis points on average, while Turquoise and Bats are quoting worse prices, at 7.19 and 10.10 basis points, respectively. But there is a large difference across stocks. For HFrag stocks both Turquoise and Bats seem to have quite competitive quotes. The relative amount of depth quoted at the best prices (not necessarily depth quoted at the same prices) is generally the highest on Euronext, with on average 40% of the total amount of shares offered, followed by Chi-X (23%), and then Bats (19%) and Turquoise (18%). What is remarkable is that for HFrag stocks on average only 14% and 15% of depth is provided on Turquoise and Bats respectively, while for LFrag stocks this amounts to 21% on Turquoise and 26% on Bats. One reason is that the quotes from Turquoise and Bats for LFrag stocks are far less competitive, and thus this depth is situated deeper into the consolidated book.

2.5 Results

2.5.1 Return Predictability on the Own Venue

Table 2.2 presents a summary of results for Equation (2.1). The table shows (by stock) the mean coefficient of the lagged order book imbalance, the mean t-statistic (based on Newey-West standard errors) within parentheses and the fraction of estimated coefficients that is significantly positive or negative within square brackets. In general, coefficients are largely positive and significant. On the consolidated market level on average 99 percent of coefficients are significantly positive, with a mean t-statistic of 6.81 and adjusted R^2 of 1.85 percent. Returns of HFrag stocks are on average more predictable, with a mean t-statistic of 8.35 and adjusted R^2 of 2.55 percent; consolidated market returns of LFrag stocks are the least predictable, with a mean t-statistic of 4.76 and adjusted R^2 of 1.01 percent. As Table 2.1 shows, both stock subsamples do not only differ

Table 2.2: Return Predictability on the Own Venue

This table shows results from the time series Equation (2.1) for the consolidated market, as well as the four different trading venues: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). The dependent variable is the 10 second midquote return on the consolidated market or the trading venue, and the independent variable is the lagged order book imbalance on the consolidated market or the venue. We estimate the equation on a daily basis for 30 stocks. Each line shows the average coefficient and the average t-statistic (based on Newey-West standard errors) of the daily estimates, followed by the fraction of days for which the lagged order book imbalance is significantly positive or significantly negative at the 5% level for that particular stock, and finally the average adjusted R^2. Coefficients of order book imbalances are multiplied by 10,000 for readability. The last six lines of the table show the cross-sectional averages and medians for all stocks (All), the five stocks that have the highest fragmentation (HFrag), and the five stocks that have the lowest fragmentation (LFrag). *, **, *** denote significance at the 10%, 5% and 1% level, respectively, based on the average t-statistic. †, ††, † † † indicates that more than 20 percent, 50 percent or 90 percent of stock-days have either significant positive, or negative coefficients, at the 5% level.

$$r_t^i = \alpha + \beta^i b_{t-1}^i + u_t^i$$

Ticker	$r_t^{Con} = \alpha + \beta^{Con} b_{t-1}^{Con} + u_t^{Con}$					$r_t^{Enx} = \alpha + \beta^{Enx} b_{t-1}^{Enx} + u_t^{Enx}$				
	b_{t-1}^{Con}					b_{t-1}^{Enx}				
	coeff				$\bar{R}^2$	coeff				$\bar{R}^2$
BNPP	0.78***	(6.61)	[1.00]†††	[0.00]	1.52	0.76***	(5.95)	[0.98]†††	[0.00]	1.24
EDF	0.47***	(6.50)	[1.00]†††	[0.00]	1.55	0.64***	(7.69)	[1.00]†††	[0.00]	2.20
LAGA	0.69***	(8.10)	[1.00]†††	[0.00]	2.67	0.72***	(7.43)	[1.00]†††	[0.00]	2.20
LVMH	0.60***	(6.72)	[1.00]†††	[0.00]	1.68	0.73***	(7.12)	[1.00]†††	[0.00]	1.90
SASY	0.62***	(8.13)	[1.00]†††	[0.00]	2.39	0.63***	(7.37)	[1.00]†††	[0.00]	1.95
SEVI	0.71***	(9.27)	[1.00]†††	[0.00]	3.49	0.65***	(8.06)	[1.00]†††	[0.00]	2.55
SOGN	0.65***	(4.96)	[0.97]†††	[0.00]	0.87	0.81***	(5.46)	[0.99]†††	[0.00]	1.05
STM	0.53***	(6.07)	[1.00]†††	[0.00]	1.28	0.60***	(5.78)	[1.00]†††	[0.00]	1.26
TCFP	0.48***	(5.87)	[1.00]†††	[0.00]	1.41	0.57***	(6.05)	[1.00]†††	[0.00]	1.49
TOTF	0.55***	(6.69)	[0.99]†††	[0.00]	1.51	0.56***	(6.03)	[0.99]†††	[0.00]	1.22
COR	0.45***	(5.96)	[0.99]†††	[0.00]	1.40	0.56***	(6.34)	[0.98]†††	[0.00]	1.69
ELSN	0.44***	(7.07)	[1.00]†††	[0.00]	1.83	0.46***	(6.36)	[1.00]†††	[0.00]	1.55
FUGR	0.51***	(6.07)	[1.00]†††	[0.00]	1.44	0.61***	(6.48)	[1.00]†††	[0.00]	1.75
HEIN	0.52***	(7.40)	[1.00]†††	[0.00]	2.09	0.52***	(6.58)	[1.00]†††	[0.00]	1.70
ING	0.77***	(5.36)	[0.97]†††	[0.00]	1.07	0.80***	(5.06)	[0.98]†††	[0.00]	0.91
ISPA	0.78***	(6.48)	[0.99]†††	[0.00]	1.43	0.73***	(5.60)	[0.99]†††	[0.00]	1.05
PHG	0.85***	(8.76)	[1.00]†††	[0.00]	2.71	0.81***	(7.77)	[1.00]†††	[0.00]	2.09
RDSa	0.67***	(10.86)	[1.00]†††	[0.00]	4.08	0.65***	(9.46)	[1.00]†††	[0.00]	3.13
SBMO	0.83***	(8.51)	[1.00]†††	[0.00]	2.90	0.84***	(7.84)	[1.00]†††	[0.00]	2.44
UN	0.75***	(10.61)	[1.00]†††	[0.00]	3.94	0.70***	(9.13)	[1.00]†††	[0.00]	2.92
ABI	0.66***	(6.86)	[1.00]†††	[0.00]	1.69	0.57***	(5.31)	[1.00]†††	[0.00]	1.01
ACKB	0.46***	(3.99)	[0.96]†††	[0.00]	0.63	0.42***	(3.26)	[0.86]††	[0.00]	0.43
BEFB	0.32***	(3.35)	[0.91]†††	[0.00]	0.45	0.36***	(3.09)	[0.85]††	[0.00]	0.39
BEKB	0.72***	(6.56)	[1.00]†††	[0.00]	1.81	0.68***	(5.57)	[0.99]†††	[0.00]	1.32
FOR	0.99***	(7.20)	[1.00]†††	[0.00]	1.95	0.82***	(5.57)	[1.00]†††	[0.00]	1.11
KBC	1.06***	(6.88)	[1.00]†††	[0.00]	1.83	0.85***	(4.96)	[1.00]†††	[0.00]	0.93
MSTAR	0.45***	(6.09)	[1.00]†††	[0.00]	1.49	0.47***	(5.47)	[1.00]†††	[0.00]	1.24
NYR	0.86***	(6.01)	[1.00]†††	[0.00]	1.52	1.07***	(6.72)	[1.00]†††	[0.00]	1.93
OMEP	0.36***	(3.49)	[0.91]†††	[0.00]	0.48	0.35***	(3.00)	[0.83]††	[0.00]	0.36
UCB	0.61***	(6.75)	[1.00]†††	[0.00]	1.88	0.57***	(5.52)	[0.99]†††	[0.00]	1.29
Mean All	0.64***	(6.81)	[0.99]†††	[0.00]	1.85	0.65***	(6.23)	[0.98]†††	[0.00]	1.56
Median All	0.60***	(6.69)			1.60	0.61***	(6.20)			1.38
Mean HF	0.66***	(8.35)	[1.00]†††	[0.00]	2.55	0.65***	(7.40)	[1.00]†††	[0.00]	1.99
Median HF	0.63***	(8.32)			2.38	0.62***	(7.43)			1.86
Mean LF	0.56***	(4.76)	[0.96]†††	[0.00]	1.01	0.59***	(4.41)	[0.91]†††	[0.00]	0.92
Median LF	0.48***	(4.52)			0.76	0.47***	(4.09)			0.61

Table 2.2 continued.

$r_t^i = \alpha + \beta^i b_{t-1}^i + u_t^i$

Ticker	$r_t^{Chi} = \alpha + \beta^{Chi} b_{t-1}^{Chi} + u_t^{Chi}$					$r_t^{Tur} = \alpha + \beta^{Tur} b_{t-1}^{Tur} + u_t^{Tur}$				
	b_{t-1}^{Chi}					b_{t-1}^{Tur}				
	coeff				$\bar{R}^2$	coeff				$\bar{R}^2$
BNPP	0.84***	(5.99)	[1.00]† † †	[0.00]	1.21	5.14***	(4.48)	[0.87]††	[0.00]	0.86
EDF	0.48***	(4.94)	[0.96]† † †	[0.00]	0.84	0.48**	(2.54)	[0.57]††	[0.00]	0.28
LAGA	0.63***	(5.55)	[0.97]† † †	[0.00]	1.18	0.61***	(3.96)	[0.84]††	[0.00]	0.56
LVMH	0.55***	(4.87)	[0.95]† † †	[0.00]	0.87	0.90***	(3.28)	[0.71]††	[0.01]	0.56
SASY	0.61***	(6.18)	[0.96]† † †	[0.00]	1.43	0.50***	(4.21)	[0.79]††	[0.00]	0.70
SEVI	0.71***	(6.18)	[0.95]† † †	[0.00]	1.53	0.99***	(3.75)	[0.74]††	[0.00]	0.57
SOGN	0.74***	(4.08)	[0.92]† † †	[0.00]	0.59	0.82***	(3.08)	[0.76]††	[0.00]	0.35
STM	0.54***	(4.79)	[0.96]† † †	[0.00]	0.80	0.78***	(5.52)	[0.95]† † †	[0.00]	1.06
TCFP	0.45***	(3.77)	[0.94]† † †	[0.00]	0.52	0.46***	(3.51)	[0.82]††	[0.01]	0.42
TOTF	0.55***	(5.64)	[0.98]† † †	[0.00]	1.06	0.66***	(5.02)	[0.89]††	[0.00]	0.88
COR	0.42***	(4.08)	[0.92]† † †	[0.00]	0.62	0.42***	(3.95)	[0.91]† † †	[0.00]	0.56
ELSN	0.47***	(4.44)	[0.93]† † †	[0.00]	0.72	0.36***	(3.83)	[0.92]† † †	[0.00]	0.47
FUGR	0.53***	(3.74)	[0.88]††	[0.00]	0.58	0.41***	(3.52)	[0.81]††	[0.00]	0.51
HEIN	0.45***	(5.30)	[0.97]† † †	[0.00]	1.03	0.45***	(3.91)	[0.85]††	[0.00]	0.56
ING	0.77***	(4.52)	[0.94]† † †	[0.00]	0.73	0.86***	(3.77)	[0.82]††	[0.00]	0.56
ISPA	0.76***	(5.27)	[0.99]† † †	[0.00]	0.94	0.88***	(3.98)	[0.78]††	[0.01]	0.66
PHG	0.76***	(7.05)	[1.00]† † †	[0.00]	1.68	0.75***	(5.48)	[0.91]† † †	[0.00]	1.05
RDSa	0.64***	(9.60)	[1.00]† † †	[0.00]	3.06	0.55***	(6.63)	[0.97]† † †	[0.00]	1.40
SBMO	0.81***	(5.74)	[0.95]† † †	[0.00]	1.21	0.74***	(4.40)	[0.88]††	[0.00]	0.72
UN	0.65***	(7.70)	[1.00]† † †	[0.00]	2.13	0.63***	(6.21)	[0.93]† † †	[0.00]	1.38
ABI	0.70***	(5.88)	[0.99]† † †	[0.00]	1.22	1.23***	(4.06)	[0.87]††	[0.00]	0.68
ACKB	2.32***	(4.57)	[0.98]† † †	[0.00]	1.26	1.17***	(3.76)	[0.80]††	[0.02]	1.40
BEFB	1.48***	(4.91)	[0.92]† † †	[0.00]	1.30	1.12***	(3.24)	[0.76]††	[0.06]	1.27
BEKB	1.14***	(4.97)	[0.96]† † †	[0.00]	0.98	1.22***	(4.88)	[0.84]††	[0.04]	1.81
FOR	1.12***	(4.75)	[0.91]† † †	[0.00]	0.81	-1.10	(1.55)	[0.38]†	[0.08]	0.27
KBC	1.60***	(4.73)	[0.86]††	[0.00]	0.87	1.80***	(3.56)	[0.86]††	[0.00]	0.44
MSTAR	0.52***	(3.62)	[0.89]††	[0.00]	0.53	1.07***	(3.01)	[0.71]††	[0.05]	0.68
NYR	1.31***	(3.95)	[0.81]††	[0.00]	0.74	1.17***	(2.59)	[0.55]††	[0.03]	0.54
OMEP	1.78***	(2.87)	[0.75]††	[0.00]	0.53	0.74***	(2.61)	[0.64]††	[0.01]	0.43
UCB	0.61***	(4.47)	[0.96]† † †	[0.00]	0.76	0.86***	(3.62)	[0.78]††	[0.06]	0.85
Mean All	0.82***	(5.15)	[0.94]† † †	[0.00]	1.06	0.80***	(3.94)	[0.80]††	[0.01]	0.75
Median All	0.61***	(5.04)			0.87	0.61***	(4.02)			0.48
Mean HF	0.61***	(6.39)	[0.98]† † †	[0.00]	1.47	0.60***	(4.97)	[0.88]††	[0.00]	0.92
Median HF	0.58***	(6.48)			1.38	0.55***	(5.13)			0.75
Mean LF	1.59***	(4.25)	[0.88]††	[0.00]	0.95	1.09***	(3.43)	[0.72]††	[0.03]	1.09
Median LF	1.03***	(4.25)			0.75	0.92***	(3.58)			0.52

Table 2.2 continued.

	$r_t^i = \alpha + \beta^i b_{t-1}^i + u_t^i$				
Ticker	$r_t^{Bat} = \alpha + \beta^{Bat} b_{t-1}^{Bat} + u_t^{Bat}$				
	b_{t-1}^{Bat}				
	coeff				$\bar{R}^2$
BNPP	1.54***	(6.03)	[0.98]† † †	[0.00]	1.57
EDF	1.04***	(5.69)	[0.92]† † †	[0.00]	1.40
LAGA	1.05***	(4.42)	[0.85]††	[0.00]	0.93
LVMH	0.74***	(4.38)	[0.89]††	[0.00]	0.81
SASY	0.80***	(5.07)	[0.94]† † †	[0.00]	1.07
SEVI	0.76***	(4.50)	[0.89]††	[0.00]	0.92
SOGN	1.53***	(3.73)	[0.84]††	[0.00]	0.57
STM	0.64***	(5.11)	[0.94]† † †	[0.00]	1.01
TCFP	0.53***	(3.31)	[0.87]††	[0.00]	0.45
TOTF	0.55***	(4.86)	[0.94]† † †	[0.00]	0.91
COR	0.90***	(4.78)	[0.91]† † †	[0.00]	1.11
ELSN	0.41***	(3.76)	[0.84]††	[0.00]	0.56
FUGR	0.49***	(2.91)	[0.78]††	[0.00]	0.38
HEIN	1.02***	(4.62)	[0.83]††	[0.00]	1.21
ING	0.77***	(3.79)	[0.90]† † †	[0.00]	0.55
ISPA	0.74***	(4.83)	[0.97]† † †	[0.00]	0.78
PHG	0.61***	(5.13)	[0.95]† † †	[0.00]	0.89
RDSa	0.57***	(7.99)	[1.00]† † †	[0.00]	2.15
SBMO	1.06***	(4.74)	[0.89]††	[0.00]	0.90
UN	0.59***	(6.41)	[0.98]† † †	[0.00]	1.40
ABI	1.01***	(5.69)	[0.93]† † †	[0.00]	1.33
ACKB	5.70***	(5.10)	[0.93]† † †	[0.00]	2.27
BEFB	1.12***	(3.92)	[0.83]††	[0.02]	1.41
BEKB	2.33***	(6.04)	[0.98]† † †	[0.00]	1.97
FOR	0.95***	(3.42)	[0.68]††	[0.00]	0.56
KBC	4.00***	(7.17)	[0.98]† † †	[0.00]	2.37
MSTAR	0.45***	(4.06)	[0.72]††	[0.04]	1.19
NYR	3.08***	(4.92)	[0.85]††	[0.00]	1.46
OMEP	1.06***	(3.62)	[0.67]††	[0.03]	1.94
UCB	1.27***	(4.78)	[0.92]† † †	[0.00]	1.07
Mean All	1.24***	(4.84)	[0.89]††	[0.00]	1.17
Median All	0.73***	(4.82)			0.84
Mean HF	0.72***	(5.22)	[0.93]† † †	[0.00]	1.10
Median HF	0.59***	(5.19)			0.90
Mean LF	2.69***	(4.76)	[0.86]††	[0.01]	1.81
Median LF	1.71***	(5.06)			1.36

in fragmentation, but also have very different size, volume and liquidity. The former consists of large, liquid and actively traded stocks, while the latter are much smaller and less liquid, with the lower trading volumes and fewer quote updates. As shown in Tombeur and Wuyts (2015) more actively traded stocks have a higher return predictability at very short horizons.

When comparing predictability across venues, returns on Euronext are on average the most predictable (mean t-statistic 6.23, adjusted R^2 1.56 percent), followed by Chi-X (mean t-statistic 5.15, adjusted R^2 1.06 percent) and Bats (mean t-statistic 4.84, adjusted R^2 1.17 percent), and finally Turquoise has the least predictable returns (mean t-statistic 3.94, adjusted R^2 0.75 percent). Lagged order book imbalances seem to be the best predictors of short-term returns on the most active trading venues (Euronext and Chi-X). However, while Turquoise slightly outperforms Bats on most measures of market activity and liquidity, its returns are clearly less predictable using the state of the order book. Thus, there appears to be a ranking as to which returns are the best predictable, and this ranking is in line with the trading activity and liquidity of these trading venues. Across stocks, note that for LFrag stocks predictability is lower on all trading venues, with lower t-statistics, lower adjusted R^2 and fewer significantly positive coefficients. The result that short-term predictability is higher for more liquid stocks and more competitive trading venues is in line with predictability being driven by order choice rather than informed trading.

2.5.2 Return Predictability Across Venues

We now turn to return predictability across venues. Table 2.3 presents a summary of results of Equation (2.3). Each block of three columns shows results for a different venue (as indicated by the header) and for each venue we display results for the full sample (All), the high fragmentation stock sample (HFrag), and the low fragmentation stock sample (LFrag). For each independent variable we show the mean coefficient estimate, the mean t-statistic (within parentheses), the fraction of stock-days for which the coefficient is significantly positive, and the fraction of stock-days for which it is significantly negative (within square brackets).

As before, own venue order book imbalances can predict returns, while the order book imbalance from the consolidated competitors only mildly contributes to return predictability. The contribution is the largest on Euronext, where the mean adjusted R^2 increases slightly by adding $b_{t-1}^{Con\backslash i}$, from 1.56 to 1.70. Cross-venue predictability is much smaller on Chi-X, with an increase in mean adjusted R^2 from 1.06 to 1.13. For Turquoise and Bats the contribution of competing venues to return predictability is almost non-existent, and mean coefficients are even negative. So similarly as for Equation (2.1), cross-venue predictability of returns ranks according to venue trading activity and liquidity. Across stocks competing venues contribute to predictability the most for HFrag stocks, for all venues. Even on Turquoise and Bats there is an indication that, for these highly fragmented and more liquid stocks, the order books of competing venues matter for future returns. For LFrag stocks the opposite is observed. On alternative venues order book imbalances from competing venues even negatively predicts returns.

Instead of aggregating the order books of competing venues into one consolidated order book, we also look at the contribution of the different venues separately by estimating Equation (2.4). Results are summarized in Table 2.4. When we compare the predictive ability for consolidated market returns across venues we observe the same pattern as before: the Euronext order book contributes the most (mean coefficient 0.43, t-statistic 4.00), followed by Chi-X (mean coefficient 0.29, t-statistic 2.73), with no significant contribution of Turquoise and Bats. Furthermore, results indicate that returns are still best predicted by own venue

Table 2.3: Return Predictability Across Venues: Order Book Imbalance

This table shows results from the time series Equation (2.3) for the four different trading venues: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). The dependent variable is the 10 second midquote return on the trading venue, and the independent variables are the lagged order book imbalance on the venue itself, and the lagged order book imbalance at the best consolidated market quotes *excluding* the venue itself. Order book imbalances are defined in Equation (2.2). We estimate the equation on a daily basis for 30 stocks. For each included venue we display results in three columns, (1) for the full sample (All), (2) for the five stocks that have the highest fragmentation (HFrag), and (3) for the five stocks that have the lowest fragmentation (LFrag). Each line shows the average coefficient and the average t-statistic (based on Newey-West standard errors) of the daily estimates, followed by the fraction of days for which the lagged order book imbalance is significantly positive or significantly negative at the 5% level for that particular stock, and finally the average adjusted R^2. Coefficients of order book imbalances are multiplied by 10,000 for readability. We suppress estimates for the constant α. *, **, *** denote significance at the 10%, 5% and 1% level, respectively, based on the average t-statistic. †, ††, † † † indicates that more than 20 percent, 50 percent or 90 percent of stock-days have either significant positive, or negative coefficients, at the 5% level.

	$r_t^i = \alpha + \beta^i b_{t-1}^i + \beta^{Con\backslash i} b_{t-1}^{Con\backslash i} + u_{i,t}$											
	r_t^{Enx}			r_t^{Chi}			r_t^{Tur}			r_t^{Bat}		
	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag
b_{t-1}^i	0.62***	0.60***	0.58***	0.82***	0.58***	1.62***	0.81***	0.60***	1.10***	1.24***	0.71***	2.73***
	(5.82)	(6.72)	(4.30)	(4.78)	(5.84)	(4.21)	(3.89)	(4.98)	(3.40)	(4.76)	(5.10)	(4.75)
	[0.98]† † †	[1.00]† † †	[0.90]† † †	[0.93]† † †	[0.98]† † †	[0.89]††	[0.79]††	[0.87]††	[0.72]††	[0.88]††	[0.93]† † †	[0.85]††
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.03]	[0.00]	[0.00]	[0.01]
$b_{t-1}^{Con\backslash i}$	0.17	0.20**	0.08	0.05	0.13	-0.15	-0.04	0.02	-0.18	-0.02	0.06	-0.26
	(1.64)	(2.51)	(0.61)	(0.97)	(1.41)	(-0.24)	(0.32)	(0.55)	(-0.38)	(0.40)	(0.75)	(-0.50)
	[0.42]†	[0.65]††	[0.16]	[0.25]†	[0.34]†	[0.06]	[0.15]	[0.21]†	[0.04]	[0.16]	[0.24]†	[0.04]
	[0.01]	[0.00]	[0.02]	[0.02]	[0.00]	[0.11]	[0.07]	[0.06]	[0.12]	[0.06]	[0.05]	[0.16]
$\bar{R}^2$	1.70	2.23	0.97	1.13	1.57	0.99	0.80	0.99	1.13	1.22	1.18	1.86

Table 2.4: Return Predictability Across Multiple Venues: Order Book Imbalance

This table shows results from the time series Equation (2.4) for the consolidated market and four different trading venues: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). The dependent variable is the 10 second midquote return on the consolidated market or the trading venue, and the independent variables are the lagged order book imbalances on all venues in the sample. Order book imbalances are defined in Equation (2.2). We estimate the equation on a daily basis for 30 stocks. For each included venue we display results in three columns, (1) for the full sample (All), (2) for the five stocks that have the highest fragmentation (HFrag), and (3) for the five stocks that have the lowest fragmentation (LFrag). Each line shows the average coefficient and the average t-statistic (based on Newey-West standard errors) of the daily estimates, followed by the fraction of days for which the lagged order book imbalance is significantly positive or significantly negative at the 5% level for that particular stock, and finally the average adjusted R^2. Coefficients of order book imbalances are multiplied by 10,000 for readability. We suppress estimates for the constant α. *, **, *** denote significance at the 10%, 5% and 1% level, respectively, based on the average t-statistic. †, ††, ††† indicates that more than 20 percent, 50 percent or 90 percent of stock-days have either significant positive, or negative coefficients, at the 5% level.

$r^i_t = \alpha + \beta^i b^i_{t-1} + \sum_{j \neq i} \beta^j b^j_{t-1} + u_{i,t}$															
	r^{Con}_t			r^{Enx}_t			r^{Chi}_t			r^{Tur}_t			r^{Bat}_t		
	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag
b^{Enx}_{t-1}	0.43***	0.39***	0.45***	0.61***	0.60***	0.56***	0.10	0.14	-0.06	-0.01	0.02	-0.15	0.00	0.07	-0.29
	(4.00)	(4.36)	(3.32)	(5.71)	(6.69)	(3.89)	(1.11)	(1.44)	(0.06)	(0.55)	(0.53)	(-0.12)	(0.48)	(0.68)	(-0.44)
	[0.89]††	[0.94]†††	[0.76]††	[0.97]†††	[1.00]†††	[0.83]††	[0.29]†	[0.36]†	[0.08]	[0.18]	[0.18]	[0.07]	[0.16]	[0.19]	[0.03]
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.02]	[0.01]	[0.07]	[0.04]	[0.03]	[0.10]	[0.05]	[0.03]	[0.13]
b^{Chi}_{t-1}	0.29***	0.33***	0.11	0.15	0.18**	0.03	0.70***	0.57***	1.21***	0.03	0.08	0.06	0.04	0.11	-0.09
	(2.73)	(3.79)	(0.69)	(1.31)	(2.04)	(0.10)	(4.84)	(5.87)	(3.86)	(0.30)	(0.68)	(-0.08)	(0.40)	(0.95)	(-0.29)
	[0.65]††	[0.88]††	[0.16]	[0.34]†	[0.54]††	[0.07]	[0.94]†††	[0.98]†††	[0.86]††	[0.12]	[0.18]	[0.05]	[0.12]	[0.22]†	[0.03]
	[0.00]	[0.00]	[0.01]	[0.01]	[0.01]	[0.04]	[0.00]	[0.00]	[0.00]	[0.05]	[0.03]	[0.07]	[0.04]	[0.01]	[0.09]
b^{Tur}_{t-1}	0.03	-0.04	0.05	0.01	-0.04	0.05	0.00	-0.02	-0.02	0.78***	0.58***	1.09***	-0.05	0.03	-0.31
	(0.13)	(-0.53)	(0.27)	(0.02)	(-0.55)	(0.25)	(0.11)	(-0.22)	(0.01)	(3.73)	(4.73)	(3.15)	(-0.11)	(0.25)	(-0.48)
	[0.12]	[0.04]	[0.14]	[0.10]	[0.05]	[0.12]	[0.08]	[0.04]	[0.07]	[0.78]††	[0.86]††	[0.70]††	[0.06]	[0.10]	[0.05]
	[0.09]	[0.15]	[0.07]	[0.09]	[0.16]	[0.06]	[0.05]	[0.06]	[0.05]	[0.01]	[0.00]	[0.04]	[0.08]	[0.04]	[0.14]
b^{Bat}_{t-1}	-0.01	0.06	-0.06	-0.03	0.03	-0.03	0.01	0.04	0.02	0.08	0.09	0.16	1.21***	0.69***	2.59***
	(-0.05)	(0.70)	(-0.38)	(-0.24)	(0.41)	(-0.26)	(-0.03)	(0.40)	(0.07)	(0.01)	(0.54)	(0.19)	(4.57)	(4.82)	(4.41)
	[0.09]	[0.20]	[0.03]	[0.06]	[0.14]	[0.04]	[0.06]	[0.10]	[0.07]	[0.08]	[0.13]	[0.10]	[0.88]††	[0.92]†††	[0.84]††
	[0.10]	[0.04]	[0.10]	[0.12]	[0.04]	[0.10]	[0.07]	[0.03]	[0.05]	[0.08]	[0.03]	[0.06]	[0.00]	[0.00]	[0.01]
$\bar{R}^2$	1.40	1.74	0.89	1.79	2.27	1.08	1.26	1.65	1.06	0.92	1.09	1.30	1.29	1.25	1.90

order book imbalances. Only for HFrag stocks on Euronext and Chi-X there is some contribution of order books across both venues. Overall, these results present weak evidence at best for Hypothesis 1 that imbalances in the state of the order book can predict returns across venues.

But Hypothesis 2 states that the relationship between the state of the order book and future returns depends on the relative position of the quotes in the price grid. We therefore decompose depth that is quoted by all competing venues into depth quoted at better prices ($Con\backslash i(B)$), the same prices ($Con\backslash i(S)$) and worse prices ($Con\backslash i(W)$) (compared to venue i prices). We regress returns on lagged classified order book imbalances, $b_t^{Con\backslash i(B)}$, $b_t^{Con\backslash i(S)}$ and $b_t^{Con\backslash i(W)}$, as indicated by Equation (2.6) and present a summary of results in Table 2.5. We find that imbalances in the state of the order book of competing venues are the strongest predictor of returns when these venues are quoting better prices than venue i. For the full stock sample mean coefficients (t-statistics) are 0.38 (3.40), 0.34 (2.42), 0.41 (2.30) and 0.40 (2.12) for Euronext, Chi-X, Turquoise and Bats respectively. An explanation for this finding is that queue-jumping from order book j to order book i is more likely when the former is very competitive. As such, switching between order books has the strongest effect on quoted prices (and thus midquote returns) under these circumstances. Our findings are thus in line with Hypothesis 2.

While imbalances measured at more competitive prices strongly predict returns in the direction of the imbalance, imbalances obtained from worse prices predict returns in the opposite direction, i.e. reversals. Relatively more bid (ask) depth on less competitive venues $Con\backslash i(W)$ predicts the midquote to drop (rise) on venue i. Suppose a positive order book imbalance is observed at worse prices, which means that competing venues j are quoting relatively more bid depth behind the best prices of venue i. This provides more incentives for traders to jump the bid queue on venue i (by submitting a limit buy order on venue j) rather than vice versa. Furthermore, when competing venues are quoting relatively more bid depth at a worse price than venue i, it suggests that the bid quote on venue i is (relatively) too high, e.g., because competition for order execution on the bid side of venue i has just raised the bid price. This is in line with the model of Goettler, Parlour, and Rajan (2005) who state that, within a single order book, when the bid (ask) side is thicker behind the best quotes than at the best quotes, this can signal a midquote above the common value of the asset. The effect is the strongest on Turquoise (mean coefficient -0.74, t-statistic -2.87) and Bats (mean coefficient -0.52, t-statistic -2.35) because these venues generally do not have better prices than their competitors, but when they do, they are more likely to be out of line with fundamentals.

For order book imbalances at the same prices ($b_t^{Con\backslash i(S)}$) the sign of the effect depends on the venue. On Euronext and Chi-X the queue-jumping from order book j to order book i dominates and future returns are positively related

Table 2.5: Return Predictability Across Venues: The Effect of Quote Competitiveness

Note: This table shows results from the time series Equation (2.6) for the four different trading venues: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). The dependent variable is the 10 second midquote return on the trading venue, and the independent variables are the lagged order book imbalance on the venue itself, and lagged order book imbalances from competing venues, classified according to the relative position on the price grid: better than the quotes of venue i ($i(B)$), at the same quotes as those of venue i ($i(S)$), and at worse quotes than those of venue i ($i(W)$). Order book imbalances are defined in Equations (2.2) and (2.5). We estimate the equation on a daily basis for 30 stocks. For each included venue we display results in three columns, (1) for the full sample (All), (2) for the five stocks that have the highest fragmentation (HFrag), and (3) for the five stocks that have the lowest fragmentation (LFrag). Each line shows the average coefficient and the average t-statistic (based on Newey-West standard errors) of the daily estimates, followed by the fraction of days for which the lagged order book imbalance is significantly positive or significantly negative at the 5% level for that particular stock, and finally the average adjusted R^2. Coefficients of order book imbalances are multiplied by 10,000 for readability. We suppress estimates for the constant α. *, **, *** denote significance at the 10%, 5% and 1% level, respectively, based on the average t-statistic. †, ††, † † † indicates that more than 20 percent, 50 percent or 90 percent of stock-days have either significant positive, or negative coefficients, at the 5% level.

$$r_t^i = \alpha + \beta^i b_{t-1}^i + \beta^{Con\setminus i(B)} b_{t-1}^{Con\setminus i(B)} + \beta^{Con\setminus i(S)} b_{t-1}^{Con\setminus i(S)} + \beta^{Con\setminus i(W)} b_{t-1}^{Con\setminus i(W)} + u_{i,t}$$

	r_t^{Enx}			r_t^{Chi}			r_t^{Tur}			r_t^{Bat}		
	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag
b_{t-1}^i	0.43***	0.38***	0.47***	0.72***	0.47***	1.35***	0.52***	0.33***	0.75**	0.82***	0.54***	1.88***
	(3.75)	(3.85)	(3.41)	(3.86)	(4.44)	(3.13)	(2.65)	(3.00)	(2.36)	(3.08)	(3.71)	(2.85)
	[0.86]††	[0.90]††	[0.77]††	[0.87]††	[0.94]† † †	[0.78]††	[0.67]††	[0.73]††	[0.62]††	[0.74]††	[0.86]††	[0.69]††
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.03]	[0.01]	[0.00]	[0.02]
$b_{t-1}^{Con\setminus i(B)}$	0.38***	0.36***	0.25*	0.34**	0.29***	0.65***	0.41**	0.38***	0.54**	0.40**	0.35***	0.45
	(3.40)	(3.61)	(1.95)	(2.42)	(2.74)	(2.78)	(2.30)	(2.94)	(2.19)	(2.12)	(2.92)	(1.35)
	[0.80]††	[0.89]††	[0.50]††	[0.60]††	[0.70]††	[0.65]††	[0.56]††	[0.70]††	[0.56]††	[0.53]††	[0.72]††	[0.34]†
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.02]
$b_{t-1}^{Con\setminus i(S)}$	0.28***	0.29***	0.14	0.02	0.13	-0.12	-0.45	-0.12	-0.79	-0.38	-0.10	-1.06**
	(2.95)	(3.46)	(1.16)	(1.05)	(1.33)	(-0.04)	(-0.79)	(-0.34)	(-1.50)	(-0.80)	(-0.18)	(-2.22)
	[0.75]††	[0.88]††	[0.29]†	[0.26]†	[0.31]†	[0.10]	[0.05]	[0.05]	[0.02]	[0.06]	[0.07]	[0.00]
	[0.00]	[0.00]	[0.01]	[0.02]	[0.00]	[0.10]	[0.23]†	[0.11]	[0.39]†	[0.24]†	[0.14]	[0.53]††
$b_{t-1}^{Con\setminus i(W)}$	-0.13	-0.13*	-0.11	-0.21	-0.14*	-0.48	-0.74***	-0.51***	-0.91**	-0.52**	-0.36***	-0.61
	(-1.34)	(-1.83)	(-0.82)	(-1.12)	(-1.74)	(-1.33)	(-2.87)	(-3.36)	(-2.13)	(-2.35)	(-3.17)	(-1.10)
	[0.01]	[0.00]	[0.01]	[0.01]	[0.00]	[0.01]	[0.00]	[0.00]	[0.01]	[0.01]	[0.00]	[0.03]
	[0.31]†	[0.46]†	[0.17]	[0.26]†	[0.43]†	[0.32]†	[0.70]††	[0.79]††	[0.56]††	[0.59]††	[0.80]††	[0.30]†
$\bar{R}^2$	2.85	3.55	1.42	1.63	2.13	1.72	2.02	2.56	2.24	2.26	2.48	2.78

to the order book imbalances from the consolidated competitors $Con\backslash i(S)$. For Turquoise and Bats the relative over- or undervaluation that is implied by the relative position of the quotes matters most, and the the order book imbalance on competing venues $b_t^{Con\backslash i(S)}$ is negatively related to future returns. Overall, the size of the effect is smaller for imbalances at the same prices.

The fact that the relation between order book imbalances and future returns depends on the relative position of the quotes provides an explanation for the weak results from Tables 2.3 and 2.4. In order to separate the effect of order book imbalances from the effect of the relative positions of the quotes on the price grid, we introduce a new measure to capture the latter: the distance d_t^j between the midquote of venue i and venue j. Results of Equation (2.8) are presented in Table 2.6 and show that the distance between the midquote of venue i and the midquote of the consolidated competitors $Con\backslash i$ has a large predictive power for future venue returns. It is strongly significant for all venues and the mean adjusted R^2 increases from 1.70 to 3.18, from 1.13 to 9.87, from 0.80 tot 15.75, and from 1.22 to 10.60 for returns on Euronext, Chi-X, Turquoise and Bats respectively. This is in line with Hypothesis 3. Furthermore, the coefficient of lagged order book imbalance of the consolidated competitors $Con\backslash i$ is now larger and more significant. In fact, for Chi-X, Turquoise and Bats competing venues seem to contribute more to predictability than the own venue imbalance, and especially the lagged distance has quite some explanatory power for returns. These results suggest that Hypothesis 1 holds when we control for the relative positions of the quotes.

The large adjusted R^2 for returns on alternative venues (primarily Turquoise and Bats), compared to a relatively low adjusted R^2 for Euronext, the listing exchange, indicate that price discovery is taking place mostly on the latter venue. New information is first impounded into prices on Euronext, and later prices on alternative venues tend to adjust.

In Table 2.7 we present a summary of results of Equation (2.9) and explore return predictability within and across venues into more detail by disentangling the contribution of the different venues. Order book imbalances and distances from Euronext and Chi-X significantly predict returns, on the consolidated market as well as on separate trading venues, although for LFrag stocks their contribution is lower. By contrast, Bats and Turquoise contribute to a much lesser extent; for LFrag stocks their contribution is almost non-existent. Even for own venue returns the predictive ability of Turquoise and Bats order book imbalances is weaker than that of Euronext and Chi-X. In line with results of Equation (2.8), the adjusted R^2s for the alternative venues are much higher compared to models that only include order book imbalances.

Overall, we find that returns on a venue i can be predicted by the lagged state of the order book of venue j and thus we accept Hypothesis 1. Moreover, also the relative position of the quotes across venues contains information that

Table 2.6: Return Predictability Across Venues: Order Book Imbalance and Distance

This table shows results from the time series Equation (2.8) for the four different trading venues: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). The dependent variable is the 10 second midquote return on the trading venue, and the independent variables are the lagged order book imbalance on the venue itself, the lagged order book imbalance at the best consolidated market quotes *excluding* the venue itself, and the lagged distance between the midquotes of both order books. Order book imbalances are defined in Equation (2.2) and distance is defined in Equation (2.7). We estimate the equation on a daily basis for 30 stocks. For each included venue we display results in three columns, (1) for the full sample (All), (2) for the five stocks that have the highest fragmentation (HFrag), and (3) for the five stocks that have the lowest fragmentation (LFrag). Each line shows the average coefficient and the average t-statistic (based on Newey-West standard errors) of the daily estimates, followed by the fraction of days for which the lagged order book imbalance is significantly positive or significantly negative at the 5% level for that particular stock, and finally the average adjusted R^2. Coefficients of order book imbalances are multiplied by 10,000 for readability. We suppress estimates for the constant α. *, **, *** denote significance at the 10%, 5% and 1% level, respectively, based on the average t-statistic. †, ††, † † † indicates that more than 20 percent, 50 percent or 90 percent of stock-days have either significant positive, or negative coefficients, at the 5% level.

$$r_t^i = \alpha + \beta^i b_{t-1}^i + \beta^{Con\backslash i} b_{t-1}^{Con\backslash i} + \delta^{Con\backslash i} d_{t-1}^{Con\backslash i} + u_t^i$$

	r_t^{Enx}			r_t^{Chi}			r_t^{Tur}			r_t^{Bat}		
	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag	All	HFrag	LFrag
b_{t-1}^i	0.44***	0.40***	0.46***	0.33***	0.37***	0.06	0.15	0.05	0.43	0.18	0.23**	0.48
	(3.81)	(3.99)	(3.32)	(2.72)	(3.79)	(0.73)	(0.92)	(0.70)	(1.35)	(1.09)	(2.00)	(1.18)
	[0.85]††	[0.88]††	[0.76]††	[0.65]††	[0.83]††	[0.28]†	[0.33]†	[0.32]†	[0.40]†	[0.37]†	[0.55]††	[0.38]†
	[0.00]	[0.00]	[0.00]	[0.02]	[0.01]	[0.07]	[0.09]	[0.12]	[0.05]	[0.08]	[0.03]	[0.08]
$b_{t-1}^{Con\backslash i}$	0.31***	0.34***	0.21	0.53***	0.41***	0.93**	0.64***	0.60***	0.59	0.53***	0.50***	0.52
	(2.82)	(3.86)	(1.53)	(3.05)	(3.64)	(2.49)	(3.22)	(4.39)	(1.63)	(2.76)	(3.79)	(1.39)
	[0.69]††	[0.90]††	[0.37]†	[0.74]††	[0.84]††	[0.64]††	[0.71]††	[0.86]††	[0.44]†	[0.65]††	[0.83]††	[0.36]†
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.02]	[0.01]	[0.00]	[0.03]
$d_{t-1}^{Con\backslash i}$	0.21***	0.31***	0.04***	0.31***	0.39***	0.19***	0.42***	0.53***	0.15***	0.33***	0.43***	0.12***
	(3.75)	(4.23)	(2.59)	(13.52)	(8.81)	(4.55)	(15.61)	(19.63)	(5.39)	(11.84)	(27.01)	(5.13)
	[0.77]††	[0.84]††	[0.66]††	[0.86]††	[0.89]††	[0.94]††	[0.96]††	[0.98]††	[0.93]††	[0.95]††	[0.97]††	[0.92]††
						†	†	†	†	†	†	†
	[0.01]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
$\bar{R}^2$	3.18	3.88	1.68	9.87	10.15	8.91	15.75	19.83	6.60	10.60	12.39	5.95

Table 2.7: Return Predictability Across Multiple Venues: Order Book Imbalance and Distance

This table shows results from the time series Equation (2.9) for the consolidated market and four different trading venues: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). The dependent variable is the 10 second midquote return on the consolidated market or the trading venue, and the independent variables are the lagged order book imbalances on all venues in the sample, as well as the distance to the midquotes of the other venues. Order book imbalances are defined in Equation (2.2) and distance is defined in Equation (2.7). We estimate the equation on a daily basis for 30 stocks. For each included venue we display results in three columns, (1) for the full sample (All), (2) for the five stocks that have the highest fragmentation (HFrag), and (3) for the five stocks that have the lowest fragmentation (LFrag). Each line shows the average coefficient and the average t-statistic (based on Newey-West standard errors) of the daily estimates, followed by the fraction of days for which the lagged order book imbalance is significantly positive or significantly negative at the 5% level for that particular stock, and finally the average adjusted R^2. Coefficients of order book imbalances are multiplied by 10,000 for readability. We suppress estimates for the constant α. *, **, *** denote significance at the 10%, 5% and 1% level, respectively, based on the average t-statistic. †, ††, † † † indicates that more than 20 percent, 50 percent or 90 percent of stock-days have either significant positive, or negative coefficients, at the 5% level.

$$r_t^i = \alpha + \beta^i b_{t-1}^i + \beta^j b_{t-1}^j + \delta^j d_{t-1}^j + \mu^j ms_{t-1,\tau}^j + \gamma^j b_{t-1}^j * ms_{t-1,\tau}^j + \lambda^j d_{t-1}^j * ms_{t-1,\tau}^j + u_t^i$$

	r_t^{Con}			r_t^{Enx}			r_t^{Chi}			r_t^{Tur}			r_t^{Bat}		
	All	HF	LF	All	HF	LF	All	HF	LF	All	HF	LF	All	HF	LF
	0.31***	0.30***	0.34**	0.37***	0.35***	0.38**	0.34**	0.30***	0.43*	0.34*	0.30**	0.29	0.25	0.30**	0.05
b_{t-1}^{Enx}	(2.68)	(3.00)	(2.37)	(3.13)	(3.49)	(2.54)	(2.32)	(2.59)	(1.66)	(1.73)	(2.06)	(0.94)	(1.44)	(1.99)	(0.40)
	[0.66]††	[0.74]††	[0.57]††	[0.74]††	[0.81]††	[0.60]††	[0.59]††	[0.64]††	[0.42]†	[0.44]†	[0.51]††	[0.27]†	[0.37]†	[0.48]†	[0.15]
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]	[0.01]	[0.04]	[0.02]	[0.01]	[0.06]
	0.26**	0.27***	0.15	0.23*	0.24**	0.16	0.33**	0.33***	0.13	0.39*	0.34**	0.39	0.36*	0.31**	0.44
b_{t-1}^{Chi}	(2.16)	(2.73)	(0.86)	(1.90)	(2.41)	(0.82)	(2.56)	(3.30)	(0.76)	(1.88)	(2.52)	(0.96)	(1.72)	(2.34)	(0.81)
	[0.55]††	[0.72]††	[0.19]	[0.48]†	[0.65]††	[0.17]	[0.63]††	[0.79]††	[0.23]†	[0.48]†	[0.65]††	[0.24]†	[0.44]†	[0.61]††	[0.21]†
	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.01]	[0.01]	[0.01]	[0.05]	[0.01]	[0.01]	[0.02]	[0.01]	[0.00]	[0.02]
	0.07	0.04	0.08	0.06	0.03	0.07	0.07	0.03	0.16	0.09	0.03	0.28	0.16	0.01	0.62
b_{t-1}^{Tur}	(0.50)	(0.36)	(0.38)	(0.40)	(0.23)	(0.32)	(0.42)	(0.25)	(0.46)	(0.54)	(0.50)	(0.70)	(0.37)	(0.12)	(0.84)
	[0.13]	[0.12]	[0.07]	[0.11]	[0.10]	[0.07]	[0.14]	[0.11]	[0.15]	[0.25]†	[0.27]†	[0.25]†	[0.14]	[0.10]	[0.24]†
	[0.03]	[0.04]	[0.02]	[0.03]	[0.05]	[0.02]	[0.04]	[0.06]	[0.04]	[0.10]	[0.12]	[0.08]	[0.06]	[0.08]	[0.03]
	0.07	0.10	0.08	0.08	0.10	0.12	0.08	0.10	0.15	0.15	0.11	0.31	0.09	0.18	0.17
b_{t-1}^{Bat}	(0.54)	(1.11)	(0.29)	(0.54)	(1.06)	(0.45)	(0.50)	(1.03)	(0.34)	(0.44)	(0.84)	(0.47)	(0.72)	(1.60)	(0.62)
	[0.14]	[0.26]†	[0.10]	[0.14]	[0.25]†	[0.12]	[0.14]	[0.25]†	[0.12]	[0.15]	[0.22]†	[0.17]	[0.28]†	[0.45]†	[0.27]†
	[0.03]	[0.01]	[0.03]	[0.03]	[0.01]	[0.02]	[0.04]	[0.02]	[0.06]	[0.05]	[0.03]	[0.06]	[0.09]	[0.04]	[0.10]

d_{t-1}^{Enx}	0.02**	0.02**	0.01**				0.02***	0.03***	0.01***	0.01***	0.02***	0.01**	0.01**	0.02***	0.00*
	(2.51)	(2.54)	(2.29)				(4.20)	(4.75)	(3.78)	(2.61)	(2.60)	(2.40)	(2.19)	(2.79)	(1.81)
	[0.64]††	[0.63]††	[0.62]††				[0.71]††	[0.70]††	[0.84]††	[0.43]†	[0.36]†	[0.56]††	[0.36]†	[0.39]†	[0.43]†
	[0.00]	[0.00]	[0.00]				[0.00]	[0.00]	[0.00]	[0.02]	[0.02]	[0.01]	[0.02]	[0.02]	[0.01]
d_{t-1}^{Chi}	0.19***	0.23***	0.03*	0.17***	0.22***	0.03*				0.23**	0.27**	0.07*	0.20**	0.23**	0.06
	(2.75)	(2.64)	(1.71)	(2.99)	(2.84)	(1.83)				(2.44)	(2.53)	(1.75)	(2.08)	(2.31)	(1.25)
	[0.68]††	[0.68]††	[0.43]†	[0.71]††	[0.70]††	[0.46]†				[0.59]††	[0.61]††	[0.46]†	[0.53]††	[0.57]††	[0.31]†
	[0.01]	[0.01]	[0.01]	[0.01]	[0.01]	[0.00]				[0.02]	[0.04]	[0.01]	[0.02]	[0.03]	[0.01]
d_{t-1}^{Tur}	0.03	0.05	0.01	0.03	0.04	0.01	0.04	0.05	0.03				0.04	0.04	0.04
	(1.12)	(1.41)	(1.14)	(0.96)	(1.21)	(1.01)	(1.17)	(1.50)	(1.29)				(1.04)	(1.29)	(1.23)
	[0.27]†	[0.35]†	[0.19]	[0.23]†	[0.30]†	[0.17]	[0.30]†	[0.36]†	[0.28]†				[0.28]†	[0.33]†	[0.31]†
	[0.04]	[0.05]	[0.01]	[0.05]	[0.05]	[0.01]	[0.05]	[0.07]	[0.01]				[0.05]	[0.07]	[0.01]
d_{t-1}^{Bat}	0.05	0.08*	0.01	0.05	0.08*	0.01	0.06	0.09*	0.02	0.09	0.12*	0.04			
	(1.37)	(1.65)	(0.82)	(1.29)	(1.65)	(0.78)	(1.51)	(1.77)	(1.11)	(1.50)	(1.84)	(1.42)			
	[0.34]†	[0.40]†	[0.19]	[0.31]†	[0.40]†	[0.17]	[0.37]†	[0.44]†	[0.28]†	[0.38]†	[0.44]†	[0.37]†			
	[0.02]	[0.02]	[0.01]	[0.02]	[0.03]	[0.01]	[0.03]	[0.04]	[0.02]	[0.03]	[0.05]	[0.02]			
$\bar{R}^2$	3.76	4.21	2.58	3.91	4.59	2.17	7.47	9.72	5.70	16.72	21.49	6.82	11.62	13.81	6.60

can be used to predict returns. The relation between returns on venue i and the lagged state of the order book on competing venue j is stronger when venue j is relatively more competitive, and thus traders are more likely to queue-jump to venue i. This confirms Hypothesis 2. We find also that prices adjust across trading venues as the distance between the midquotes of two trading venues can predict returns on either of these venues, which confirms Hypothesis 3. Our findings suggest that the listing exchange (Euronext) is leading in price discovery, while the small and less liquid alternative venues (Turquoise and Bats) are followers. The strongest competing venue (Chi-X) seems to be an intermediate case.

2.5.3 The Effect of Fragmentation

The previous results indicate that there is heterogeneity in cross-venue return predictability between venues, across stocks and also over time. Hypothesis 4 states that cross-venue return predictability is stronger when trading activity is more fragmented across trading venues. In particular, as a trading venue attracts more volume, this indicates that both liquidity demanders and suppliers are finding their way to the venue. In turn, this could induce also other traders to submit orders to this venue because of a liquidity externality effect (Pagano, 1989). As a venue becomes more relevant in the order choice consideration of traders, this could reinforce the mechanism that underlies return predictability across venues.

Panel A of Table 2.8 presents a summary of estimation results of Equation (2.10), with the Euronext return as the dependent variable, and order book and price information from alternative venues as independent variables. Panel B shows results for the same equation estimated with returns from alternative venues as the dependent variable, and Euronext order book information as independent variable. These equations are estimated per stock per year. Interactions are included between, on the one hand, the lagged state of the order book of venue j and the lagged distance between midquotes of venue i and venue j, and on the other hand, the relative market share of venue j (as compared to venue i). We thus allow predictability of returns using venue j order book information to vary according to the relative market share of venues j. The goal of the model is to explain the variation in predictability and it cannot be used to predict returns because it uses information unavailable at time $t-1$.

For Euronext returns we find in Panel A that the order book of alternative venues j can better predict returns on days when venue j has a relatively larger market share. For Chi-X the mean interaction with relative market share is 0.49 (t-statistic 2.99), for Turquoise it is 1.32 (t-statistic 3.98), and for Bats it is 1.02 (t-statistic 2.62). The result holds both for HFrag stocks and LFrag stocks, and is in line with Hypothesis 4. However, we do not find that the market share affects the predictive ability of the distance between midquotes in a similar way. In fact,

Table 2.8: Return Predictability Across Multiple Venues: The Effect of Fragmentation

This table shows results from the time series Equation (2.10) for the four different trading venues: Euronext (Enx), Chi-X (Chi), Turquoise (Tur) and Bats (Bat). The dependent variable is the 10 second midquote return on the trading venue, and the independent variables are the lagged order book imbalance on the venue itself, the lagged order book imbalance at one of the competing venues, and the lagged distance between the midquotes of both order books. Order book imbalances are defined in Equation (2.2) and distance is defined in Equation (2.7). Furthermore, we include the contemporaneous daily relative market share of the competing venue and its interaction with the order book imbalance and distance. We estimate the equation on a yearly basis for 30 stocks. For each included venue we display results in three columns, (1) for the full sample (All), (2) for the five stocks that have the highest fragmentation (HFrag), and (3) for the five stocks that have the lowest fragmentation (LFrag). Each line shows the average coefficient and the average t-statistic (based on Newey-West standard errors) of the yearly estimates, followed by the fraction of stocks for which the lagged order book imbalance is significantly positive or significantly negative at the 5% level for that particular stock, and finally the average adjusted R^2. Coefficients of order book imbalances are multiplied by 10,000 for readability. We suppress estimates for the constant α. *, **, *** denote significance at the 10%, 5% and 1% level, respectively, based on the average t-statistic. †, ††, † † † indicates that more than 20 percent, 50 percent or 90 percent of stocks have either significant positive, or negative coefficients, at the 5% level. Panel A shows result with Euronext returns as dependent variable, while Panel B shows results with Chi-X, Turquoise and Bats returns as dependent variables.

	Panel A								
	$r_t^i = \alpha + \beta^i b_{t-1}^i + \beta^j b_{t-1}^j + \delta^j d_{t-1}^j + \mu^j ms_{t-1,\tau}^j + \gamma^j b_{t-1}^j * ms_{t-1,\tau}^j + \lambda^j d_{t-1}^j * ms_{t-1,\tau}^j + u_t^i$								
	r_t^{Enx}			r_t^{Enx}			r_t^{Enx}		
	All	HF	LF	All	HF	LF	All	HF	LF
b_{t-1}^i	0.56***	0.55***	0.49***	0.61***	0.62***	0.50***	0.60***	0.61***	0.51***
	(49.35)	(43.54)	(46.36)	(76.01)	(96.27)	(46.86)	(68.93)	(74.54)	(53.97)
	[1.00]† † †	[1.00]† † †	[1.00]† † †	[1.00]† † †	[1.00]† † †	[1.00]† † †	[1.00]† † †	[1.00]† † †	[1.00]† † †
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
	$j = Chi$			$j = Tur$			$j = Bat$		
b_{t-1}^j	0.07	0.05	0.05	-0.05**	-0.13***	-0.02	-0.03	-0.03	-0.01
	(1.54)	(1.17)	(1.63)	(-2.05)	(-6.58)	(-0.98)	(-1.41)	(-0.79)	(-0.64)
	[0.43]†	[0.20]	[0.40]†	[0.23]†	[0.00]	[0.20]	[0.23]†	[0.20]	[0.20]
	[0.10]	[0.00]	[0.00]	[0.47]†	[1.00]† † †	[0.20]	[0.50]†	[0.40]†	[0.40]†
d_{t-1}^j	0.07	0.12**	0.01***	0.01	0.00	0.00	0.02***	0.02**	0.00***
	(1.60)	(2.00)	(2.73)	(0.76)	(-0.04)	(1.55)	(2.64)	(2.15)	(4.64)
	[0.47]†	[0.40]†	[0.80]††	[0.30]†	[0.00]	[0.60]††	[0.53]††	[0.40]†	[0.80]††
	[0.03]	[0.00]	[0.00]	[0.20]	[0.00]	[0.20]	[0.07]	[0.00]	[0.00]
$ms_{t-1,\tau}^j$	-0.03	-0.05	-0.11	0.02	0.09	-0.28*	-0.07	-0.06	-0.18
	(-0.44)	(-0.89)	(-0.47)	(0.00)	(0.53)	(-1.67)	(-0.59)	(-0.91)	(-0.78)
	[0.03]	[0.00]	[0.00]	[0.07]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
	[0.17]	[0.20]	[0.20]	[0.07]	[0.00]	[0.20]	[0.10]	[0.20]	[0.00]
$b_{t-1}^j * ms_{t-1,\tau}^j$	0.49***	0.56***	0.38*	1.32***	1.60***	1.94***	1.02***	1.12***	1.69***
	(2.99)	(3.44)	(1.68)	(3.98)	(6.26)	(4.13)	(2.62)	(4.53)	(2.63)
	[0.57]††	[0.80]††	[0.20]	[0.67]††	[1.00]† † †	[0.80]††	[0.60]††	[0.80]††	[0.60]††
	[0.03]	[0.00]	[0.00]	[0.03]	[0.00]	[0.00]	[0.03]	[0.00]	[0.00]
$d_{t-1}^j * ms_{t-1,\tau}^j$	-0.06	-0.20	0.03	0.05	0.07	0.09	0.06	0.01	0.03
	(-0.16)	(-1.44)	(1.10)	(1.01)	(1.16)	(0.66)	(-0.22)	(-0.59)	(0.59)
	[0.17]	[0.00]	[0.40]†	[0.33]†	[0.40]†	[0.20]	[0.20]	[0.00]	[0.40]†
	[0.13]	[0.40]†	[0.00]	[0.17]	[0.00]	[0.40]†	[0.30]†	[0.40]†	[0.20]
$\bar{R}^2$	1.77	2.21	0.97	1.41	1.74	0.87	1.48	1.86	0.85

Table 2.8 continued.

Panel B

$$r_t^i = \alpha + \beta^i b_{t-1}^i + \beta^j b_{t-1}^j + \delta^j d_{t-1}^j + \mu^j ms_{t-1,\tau}^j + \gamma^j b_{t-1}^j * ms_{t-1,\tau}^j + \lambda^j d_{t-1}^j * ms_{t-1,\tau}^j + u_t^i$$

	r_t^{Ch}			r_t^{Tur}			r_t^{Bat}		
	All	HF	LF	All	HF	LF	All	HF	LF
b_{t-1}^i	0.21***	0.28***	-0.10	-0.03	0.04	-0.45	-0.08	-0.04	-0.07
	(6.08)	(7.60)	(-0.85)	(0.30)	(0.33)	(-0.71)	(0.29)	(0.04)	(0.63)
	[0.67]††	[1.00]† † †	[0.00]	[0.27]†	[0.40]†	[0.20]	[0.33]†	[0.40]†	[0.40]†
	[0.07]	[0.00]	[0.20]	[0.23]†	[0.20]	[0.40]†	[0.33]†	[0.20]	[0.40]†
	$j = Enx$			$j = Enx$			$j = Enx$		
b_{t-1}^j	1.10	0.27	3.01	1.65	1.41	-0.56	1.70	0.14	1.90
	(0.98)	(0.46)	(1.48)	(0.93)	(0.47)	(0.61)	(0.77)	(0.01)	(0.75)
	[0.23]†	[0.20]	[0.20]	[0.20]	[0.00]	[0.40]†	[0.20]	[0.00]	[0.20]
	[0.03]	[0.00]	[0.00]	[0.07]	[0.00]	[0.20]	[0.10]	[0.20]	[0.00]
d_{t-1}^j	1.15	1.10	0.81*	1.65	2.64	-0.48	1.57	1.53	0.00
	(1.38)	(1.10)	(1.91)	(1.07)	(1.18)	(-0.49)	(1.31)	(1.42)	(0.33)
	[0.30]†	[0.20]	[0.40]†	[0.33]†	[0.20]	[0.20]	[0.30]†	[0.40]†	[0.20]
	[0.03]	[0.00]	[0.00]	[0.10]	[0.00]	[0.40]†	[0.07]	[0.00]	[0.00]
$ms_{t-1,\tau}^j$	0.06	0.30***	0.05	0.53	1.04	4.71**	0.99	0.19	1.22
	(0.69)	(2.90)	(0.84)	(0.53)	(1.63)	(2.00)	(0.95)	(1.14)	(1.48)
	[0.27]†	[0.60]††	[0.40]†	[0.27]†	[0.40]†	[0.40]†	[0.40]†	[0.40]†	[0.40]†
	[0.07]	[0.00]	[0.20]	[0.13]	[0.00]	[0.20]	[0.13]	[0.00]	[0.00]
$b_{t-1}^j * ms_{t-1,\tau}^j$	-0.46	0.47	-2.44	-1.02	-0.84	1.38	-1.21	0.57	-1.69
	(-0.17)	(0.48)	(-0.91)	(-0.63)	(-0.22)	(-0.39)	(-0.47)	(0.62)	(-0.65)
	[0.07]	[0.00]	[0.00]	[0.10]	[0.00]	[0.20]	[0.10]	[0.20]	[0.00]
	[0.10]	[0.00]	[0.20]	[0.10]	[0.00]	[0.20]	[0.20]	[0.00]	[0.20]
$d_{t-1}^j * ms_{t-1,\tau}^j$	-0.85	-0.62	-0.69	-1.28	-2.25	0.69	-1.28	-1.02	0.10
	(-0.75)	(-0.37)	(-1.31)	(-0.74)	(-0.92)	(0.76)	(-0.98)	(-0.84)	(0.09)
	[0.07]	[0.00]	[0.00]	[0.10]	[0.00]	[0.40]†	[0.13]	[0.20]	[0.20]
	[0.13]	[0.00]	[0.20]	[0.17]	[0.00]	[0.20]	[0.30]†	[0.40]†	[0.20]
$\bar{R}^2$	22.10	24.89	10.95	23.09	27.78	10.42	19.37	35.67	5.07

for some stocks it appears that the predictive ability of the distance between midquotes is lower when the market share is higher.

By contrast, in Panel B we find that the predictive ability of the listing exchange is not affected by its market share. One reason that we find different results for the listing exchange and alternative venues could be due to Euronext being such a dominant venue in our sample. Even when Euronext has a relatively low market share, it is still high enough for traders to always condition their actions on the state of the Euronext order book, no matter how active alternative venues are. Whereas traders condition only on the state of the order book of alternative venues when they are relevant, i.e. when their market share is sufficiently high.

In sum, we accept Hypothesis 4 when venue i is the listing exchange and venue j is an alternative venue. We cannot accept Hypothesis 4 when we try to predict returns on an alternative venue using information from the listing exchange.

2.6 Conclusion

We extend research on high-frequency return predictability from order book information by examining whether the order book can predict returns within and across venues. We find that imbalances in the order book of one venue predict returns on another venue, and their predictive ability depends on the relative position of the order books on the price grid. Order book imbalances are stronger predictors of returns on another venue when they are obtained from relatively more competitive prices. Order book imbalances derived from worse prices may instead predict reversals, i.e. a price decline (increase) when the bid (ask) depth is relatively thicker. Because prices across venues tend to adjust to one another throughout the trading day the distance between midquotes of different order books has the most predictive power for returns. In particular, alternative venues tend to be followers rather than leading in price discovery, the more so for venues venues with a smaller market share, Turquoise and Bats, while the listing exchange, Euronext, is the leading venue. Chi-X, the largest competing venue is an intermediate case.

Alternative venues contribute more to return predictability when they have a larger market share of trading volume. Under these circumstances these venues become a relevant choice for traders to submit their orders to, and thus traders take the state of the order book on these venues into account when choosing their order. Our findings are in line with short-term predictability being driven by order choice considerations.

Chapter 3

Two Shades of Opacity: Hidden Orders versus Dark Trading

3.1 Introduction

Technology has allowed financial markets to become ever more transparent. Trading venues are now disseminating a great amount of pre- and post-trade information in real-time. This is especially the case for equities, where the electronic limit order book has become the dominant mechanism by which shares are traded.[1] This information has proved to be valuable for market participants, as it is a growing source of revenues for trading venues (Cespa and Foucault, 2014). Traders use it to update their beliefs on a security's fundamental value, but also to adjust their execution strategies (e.g., Ranaldo, 2004). Visible order submissions thus impact excution probabilities and prices in several ways. Large limit orders, for instance, provide incentives for arriving traders on the same side of the market to undercut these orders (Buti and Rindi, 2013). To the extreme, parasitic traders may engage in predatory strategies aimed at exploiting other traders' orders (Brunnermeier and Pedersen, 2005). Transparency therefore goes hand in hand with increased exposure costs for traders.

To allow traders to mitigate their exposure costs several tools have been developed that offer traders more possibilities to hide their trading intentions.

[1] Jain (2005) provides evidence that between 1977 and 2001 the leading exchanges in 101 of 120 countries investigated switched from floor-based trading to electronic trading systems, an evolution that has continued after the end of his sample period. Furthermore, the proliferation of new trading venues that have been competing with the incumbent exchanges in the U.S. and Europe has been driven by advances in information and communications technologies. Indeed, these new market places are virtual rather than physical.

These tools, which provide the possibility to submit committed orders without these being displayed to the market, can be divided into two categories. The first is that of hidden orders on lit venues, i.e. orders that are submitted to visible order books, but for which traders do not have to (fully) display the quantity they are willing to transact. These orders hide among the visible liquidity that is offered in the lit market. We refer to transactions that execute against the hidden part of the book as *hidden order trading* activity. The second tool is trading in completely dark venues. This form of dark trading does not only encompass trading in traditional dark pools (i.e. trading venues that match third-party orders without any pre-trade transparency), but also bilateral transactions executed away from the lit market. The latter includes orders that are internalized (i.e. client orders that are matched by brokers in-house) and transactions that are executed in the over-the-counter (OTC) market. We denominate these transactions away from the lit market as *dark trading* activity.[2] We use the term *opaque trading* as the general classification for all forms of trading against non-displayed orders.

Both types of opaque trading are used extensively by traders. De Winne and D'Hondt (2007) and Bessembinder, Panayides, and Venkataraman (2009) show that around 45% of the order volume submitted on Euronext in a sample of CAC40 stocks is hidden. Rosenblatt Securities argues there are more than 40 dark pools active in the U.S. with a market share of around 14% (Rosenblatt Securities, 2012) and at least 19 dark pools in Europe with an estimated market share of about 11% (Rosenblatt Securities, 2013) An even larger part of European dark volume is taking place away from regulated venues, in the OTC market, with estimates of up to 40% of volume (Gomber and Pierron, 2010). While traders have two tools to hide their trading intentions, the academic literature focuses either on hidden orders or dark venues in isolation. Nevertheless, although it is suggested that hidden orders and dark venues may compete with each other based on their similarity in allowing for opaqueness, to the best of our knowledge no prior research exists to evaluate these claims.[3] We aim to bridge this gap in the literature by exploring the interplay between hidden order trading and dark trading. Furthermore, we are the first to investigate hidden order trading explicitly in a setting where multiple venues compete for order flow. We are also the first to examine the determinants of dark trading while making a distinction between regular-sized and block-sized transaction.

Our key goal and contribution is twofold. First, we investigate the relation between hidden order trading and dark trading directly using a simultaneous equations framework. This allows us to asses whether hidden order trading and

[2]Sometimes the use of completely hidden orders on lit venues, such as hidden midpoint-pegged orders, are also referred to as a form of dark trading. In a U.S. context trading away from lit venues would be also referred to as off-exchange trading, as opposed to trades executed on the exchanges.

[3]See claims in, e.g., Buti, Rindi, and Werner (2011); Hautsch and Huang (2012); Boulatov and George (2013); Buti and Rindi (2013); Foley, Malinova, and Park (2013).

dark trading are complements or substitutes for traders. It is easy to see intuitively why hidden order trading and dark trading could be substitutes when traders decide to which venue to route their orders. Indeed, given that a trader wants to trade opaquely, if the submission of a hidden order is expected to be less profitable (e.g., due to low execution probabilities), he is likely to choose a dark order as an alternative, and vice versa. However, both types of opaque trading can be (strategic) complements due to liquidity externalities. If a trader expects that more traders will trade opaque, he will be more willing to trade dark as well. Empirically, this will result in a positive relation between hidden order volume and dark volume, which can be interpreted as both being complements.

Second, we examine how several market conditions affect both types of opaque trading. This allows us to observe which market conditions impact hidden order and dark trading similarly (and thus segment the market into visible volume and opaque volume), and which conditions have a different impact (and thus segment into hidden order volume and dark volume). Hidden order trading and dark trading can be impacted differently when certain market conditions impact venue selection rather than the order exposure decision itself. Venue selection entails the choice between a lit trading venue or a dark trading venue, and within lit venues the choice between the main listing exchange or an alternative venue. Potential determinants are selected on the basis of theoretical predictions and previous empirical research.

Understanding the interplay between both types of opaque trading is highly relevant considering the ongoing debate among regulators, practitioners and academics on the role of market fragmentation and transparency. The current proliferation of trading venues is a consequence of regulation aimed at fostering competition between trading venues. However, regulators are increasingly worried about the growing market share of dark trading and its impact on order execution quality, price discovery, market liquidity, fair access and market quality in general (see, e.g., Securities and Exchange Commission, 2010; European Commission, 2011, 2014). In a recent speech, SEC Chair Mary Jo White raises further concerns on the lack of transparency of dark venues and states that "*we must continue to examine whether dark trading volume is approaching a level that risks seriously undermining the quality of price discovery provided by lit venues*" (White, 2014). A number of empirical researchers find that high levels of dark trading indeed harm market liquidity (Degryse, de Jong, and van Kervel, 2015; Weaver, 2014; Nimalendran and Ray, 2014), price discovery (Comerton-Forde and Putniņš, 2015) or both (Hatheway, Kwan, and Zheng, 2014). However, Buti, Rindi, and Werner (2011), Gresse (2015) and Brugler (2015) find that dark trading does not harm or even increases liquidity, while Foley and Putniņš (2015) find that the effect depends on the type of dark venue.

By contrast, the use of hidden orders on lit venues is generally not associated with reduced market quality (Tuttle, 2006; Bloomfield, O'Hara, and Saar, 2015), and may even enhance liquidity under some circumstances (Aitken, Berkman, and Mak, 2001; Moinas, 2010; Boulatov and George, 2013; Buti and Rindi, 2013; Gozluklu, 2014). If hidden order trading on lit venues is indeed a substitute for dark trading, then curtailing dark trading, such as currently proposed by the European Commission for the MiFID II regulation, could bring opaque trading back to lit venues, improving market quality. If substitution from dark venues to hidden orders is less likely, then a regulatory restriction on dark trading might harm some classes of traders that now rely on dark trading venues.

For our analysis we use a high-frequency dataset (timestamped to the millisecond) spanning nearly four years, and covering all Dutch large cap index stocks. For each stock, we have information on trading volume on the main venue where the stock is listed (Euronext) and on all alternative lit trading venues where the stock is trading (Chi-X, Turquoise and Bats). Moreover, for each of these venues, we also have limit order book data. These data allow us to infer hidden order trading: we compute hidden order executions on lit venues by matching transaction and order book data on each venue. In addition, we have trades that are executed away from the lit market. These represent a collection of dark pool trades, internalized orders and OTC transactions. We distinguish these transactions into regular dark transactions and block transactions. Because we cannot explicitly distinguish between dark and block trades as in Comerton-Forde and Putniņš (2015) we use a size-based criterion, similar to Hatheway, Kwan, and Zheng (2014) and Degryse, de Jong, and van Kervel (2015).

Our main findings can be summarized as follows. First, we find that dark trading and hidden order trading are substitutes to each other: hidden order trading and dark trading negatively impact each other. We conjecture that the order submission decision is a two-stage decision. In a first-stage decision traders choose their optimal level of *exposure* in the market: how much of their order should be hidden from other market participants? In the second stage traders decide on *how* or *where* they want to hide by submitting an opaque order to a facility that provides this opportunity. This can be a dark venue, but also a hidden order on an otherwise lit venue. However, dark trading appears to be a better substitute for hidden order trading than the other way around. We find that the negative coefficient of hidden order trading is significant in the dark trading equation, but the effect of hidden order trading on dark trading is insignificant. A potential explanation lies in the observation that orders placed on dark venues are relatively more opaque than hidden orders on lit venues. Because the latter are more easily detectable their use may not be a good substitute for trading on dark venues. This finding warrants caution for regulators who want to impose restrictions on the use of dark venues. The curtailing of dark trading could be harmful for classes of traders who now make heavy use of these venues, as the lit market does not offer an adequate opaque alternative.

Secondly, we identify the effect of several market and order book conditions on hidden order trading, dark trading and block trading. We show that hidden order trading and dark trading are differently affected by total volume, the quoted spread, depth and the fraction of traders using smart order routers, while both types of opaque trading are negatively impacted by algorithmic trading.

The fraction of volume executed against hidden orders is higher on high volume days, while the opposite is true for the fraction of volume executed in dark venues. When volume is interpreted as a proxy for trading desire, the execution probability of hidden orders increases with trading volume because more traders demand immediacy and thus start trading more aggressively. Dark venues become less attractive compared to lit venues because they provide less immediacy.

Visible depth and the quoted spread are two variables that measure lit market liquidity and may affect opaque trading behavior. We find that hidden order trading is decreasing in visible depth, while dark trading is unaffected. Larger visible depth decreases the execution probability of hidden orders because visible depth has execution priority over hidden depth. The effect of visible depth on dark trading can be twofold, depending on whether liquidity demanders or liquidity providers are affected. A larger hidden depth on lit venues means liquidity demanders have fewer reasons to submit orders to dark venues, while liquidity suppliers can be tempted to substitute (hidden) limit orders on lit venues for dark orders because the lit venue is too competitive. The insignificant coefficient indicates that ultimately both effects balance each other out. A wider quoted spread diminishes hidden order trading, most likely because a wider spread makes aggressive market orders (that are more likely to trade against hidden depth) more costly. Furthermore, we find that competition for liquidity provision affects hidden order trading *across* different lit venues, as hidden order traders substitute lit venues with deeper order books for lit venues with shallower order books to increase their execution probability.

Two variables are used to capture the characteristics of the population of traders: a proxy for algorithmic trading and an estimate of the fraction of traders using smart order routers. Algorithmic trading activity negatively affects hidden order trading and dark trading, but positively impacts block trading. One explanation is that the use of algorithms substitutes for opaque orders. When algorithms are associated with predatory trading practices, large patient traders that are likely to use opaque orders may retract from both lit and dark trading venues and resort to (negotiated) block trades. Moreover, the use of algorithms also increases competition for liquidity provision on lit venues, reducing the execution probability of hidden orders. A special case of algorithms are smart order routers, which are used to tap into the liquidity of different venues at the same time. When a larger fraction of traders use smart order routers, hidden order trading diminishes, but dark trading increases. The use of smart order routers reduces the execution probability of hidden orders, while at the same time connecting more traders to dark venues.

Our research is related to two strands in the empirical literature. First, our research is related to studies that examine the use of hidden orders in limit order markets. Harris (1996), Aitken, Berkman, and Mak (2001), Tuttle (2006), De Winne and D'Hondt (2007) and Bessembinder, Panayides, and Venkataraman (2009) show that hidden orders are used to reduce the cost associated with order exposure. These studies use order-level data and focus on the submission of hidden orders to the limit order book. By contrast, we match order book level and transaction data to infer the part of hidden orders that was actually executed while it was opaque to other market participants. A similar measure of hidden order executions is used by Yao (2012) to study the information content of hidden orders. Our focus on executed volumes is more in line with the volume measures typically used to measure trading activity on dark venues. Furthermore, contrary to previous studies, we measure hidden order trading on multiple lit trading venues.

Second, our findings also relate to the empirical literature on dark trading activity. Menkveld, Yueshen, and Zhu (2015) argue that dark and lit trading venues can be ranked according to a pecking order. Midpoint matching dark venues, which have lower trading costs but offer also lower immediacy, are at the top of the pecking order, while costly immediacy providing lit trading venues are at the bottom. They show that exogenous increases in the desire for immediacy reduce the market share of dark trading venues at the bottom of the pecking order relatively more. However, they do not explicitly account for the option to submit hidden orders to lit exchanges. Buti, Rindi, and Werner (2011) and Ready (2014) both study the market share of a selection of dark venues and find evidence in line with the prediction that more lit market competition for liquidity supply drives traders to dark venues (Buti, Rindi, and Werner, 2015). Using proprietary order-level data Garvey, Huang, and Wu (2014) show that traders prefer dark venues when information asymmetry is greater. Kwan, Masulis, and McInish (2015) find that a differential regulatory treatment in the U.S. allows dark pools to attract order flow away from traditional exchanges when spreads are constrained.

Our research is also related to theoretical studies that investigate competition between lit and dark venues. Most closely related are the models of Buti, Rindi, and Werner (2015) and Brolley (2015) of competition between a dark venue and a limit order book, although both models do not incorporate hidden orders. Buti, Rindi, and Werner (2015) show that patient traders are more likely to substitute limit orders on the lit venue for dark orders when competition for liquidity provision in the limit order book is intense, i.e. when spreads are narrow and the order book is deep. A deeper order book reduces the execution probability for newly submitted limit orders, while a narrow spread makes midquote execution of a dark venue relatively more attractive. The model's results are driven by a trade-off in execution probabilities on the lit and the dark venue. Considering the priority of visible size on lit venues, the order migration to dark venues should be even stronger for hidden limit orders compared to visible limit orders. Brolley (2015) incorporates liquidity providers, asymmetric information and a trade-at rule for dark venues in his model. He shows that the size of the trade-at rule determines the type of traders that migrate to the dark venue. In turn, this affects the impact that the dark venue has on market quality and welfare.

The remainder of this chapter is organized as follows. Section 3.2 discusses the institutional background and presents the dataset employed in our analysis, while the main analysis that relates hidden order trading to dark trading is in Section 3.3. Section 3.4 further investigates the drivers of the different shades of opaque trading. Finally, Section 3.5 concludes.

3.2 Data and Sample

3.2.1 Institutional Background

Under the impulse of technological innovation and regulatory changes the financial market structure has changed from a largely consolidated system into a market where order flow is fragmented over different trading venues. In the U.S. today's equity market structure is determined by Regulation National Market System (Reg NMS) that was phased in over the course of 2007. It has accomplished that the U.S. equity market now comprises several automated trading venues that are all connected to each other. The European counterpart of Reg NMS is the Markets for Financial Instruments Directive (MiFID), which came into force in November 2007 in the European Economic Area. It has expanded the set of regulated trading venues from the Regulated Markets (RMs), which encompass the traditional national stock exchanges, to also include Multilateral Trading Facilities (MTFs) and Systematic Internalizers (SIs). Both RMs and MTFs are multilateral trading venues that match third-party order flow.[4] Most of these trading venues are also 'lit' as they are subject to pre-trade transparency requirements. They operate public limit order books that display a large amount of the orders that are submitted to their books, which makes them similar to Electronic Communication Networks (ECNs) in the U.S. However, trading venues can also be granted waivers from pre-trade transparency provided that they meet certain conditions. This allows for the existence of dark venues and hidden orders on lit trading venues in Europe.

SIs are investment firms that internalize client orders on a frequent basis and thus offer an alternative to executions on multilateral trading venues. Also SIs

[4] An important legal difference between RMs and MTFs is that the latter can be operated both by investment firms or market operators, while the former can only be operated by market operators. In practice, the most relevant difference is that RMs provide listing services, while most MTFs do not. Issuers of shares listed on RMs should comply with relevant EU regulations, while shares listed on MTFs generally fall outside the scope of these stricter regulations.

are subject to the pre-trade transparency requirements and must publish firm quotes in the stocks in which they are dealing. But because the orders from their books are not exposed to the market directly, transactions on SIs can also be considered opaque in nature. Next to the regulated trading venues, MiFID also allows transactions in the dark OTC market. As reported by Gomber and Pierron (2010) around 40% of volume is in the OTC market, particularly in broker/dealer crossing networks, which operate in a similar way as some dark venues.

Traders can submit hidden orders on all lit venues from our sample. The main trading venue (the listing exchange) is Euronext, which allows hidden liquidity only in the form of reserve orders with fixed peak size replenishments. The most important alternative trading venues, Chi-X, Bats and Turquoise, allow reserve orders as well as completely hidden limit orders and pegged hidden orders. As such, they provide traders with a broader range of options to trade opaquely in their order books.

MiFID also imposes post-trade transparency rules for shares admitted to trading on RMs: all transactions within normal trading hours have to be reported as close to real time as possible, even when they take place outside a regulated trading venue.[5] In any case, post-trade information should be reported either through facilities provided by a RM or an MTF, third party reporting facilities or proprietary arrangements.

3.2.2 Data

Data are taken from Thomson Reuters Tick History. Our sample consists of 27 Dutch large cap stocks, listed on Euronext Amsterdam. For these stocks we have a time series of 738 trading days available, which ranges from November 2007 until end September 2010. This long time window allows us to capture a variety of market and economic conditions. Moreover, the available data are very detailed since they comprise data not only of the Regulated Market where the stocks are listed (Euronext), but also include the other European lit and dark venues where these stocks are traded. The alternative lit trading venues cover the most relevant transparent limit order books for our stock sample: Chi-X, Turquoise ans Bats.[6] Chi-X is the only MTF that was operational during the full sample period. Turquoise was launched during August 2008, and starts trading most of the stocks in our sample actively on September 1, 2008; Bats was launched in October 2008 and started trading most stocks in the sample on November 14, 2008.

[5]Very large transactions can be exempted and may be reported with a delay, provided that they meet certain criteria.

[6]We exclude Nasdaq OMX Europe, Euro TLX, Virt-X, Xetra and a number of local German exchanges from our sample. Although these venues may provide depth throughout our sample period, trading is usually not on a daily basis and in any case trading volume is quite low (less than 0.5% of the sample). Moreover, not all these venues operate as a fully automated limit order markets with visible and hidden liquidity, and are therefore less relevant for our research.

The dataset has three parts. The first part contains trade and quote data, timestamped to the millisecond, of trades on all lit venues. The second part concerns order book data. These are available for each lit trading venue, also timestamped to the millisecond, identifying prices and visible depth for the ten best quotes available at each point in time. This part of the dataset is used to compute our measures for hidden order trading, as well as to identify a number of market variables, such as spreads, visible order book depth, volatility, algorithmic trading and the fraction of traders using smart order routers. We obtain information on hidden order trading by matching transaction and order book data. Upon execution we detect which part of trading volume is executed against the visible part of the order book, and which part is executed against hidden depth.

The final part of our dataset further includes transaction data for trades that are executed away from lit markets, these are so-called MiFID reported trades. Most of these trades are reported through Markit BOAT, but they can be reported through facilities provided by lit venues such as Euronext, Deutsche Börse, the London Stock Exchange (LSE) or Chi-X as well.[7] This part of the dataset allows us to compute our measure of dark trading. Do note that, in contrast to the lit venues of the dataset, we only have reported executed volume available, not the order books of these dark venues.

The fact that we use such long time series of data at a granular level uniquely distinguishes our research from other recent empirical studies into the determinants of dark trading activity, which focus on broader cross-sections of stocks (Buti, Rindi, and Werner, 2011; Ready, 2014), or a time series analysis at a more granular level (Menkveld, Yueshen, and Zhu, 2015). The most closely related to our study is Buti, Rindi, and Werner (2011), which employ one year of data from 4,482 stocks, but their sample of dark trading activity consists only of eleven dark venues who report the data voluntarily. Their sample thus excludes other types of dark trading and does not account for hidden order trading on lit venues.

We filter our data by removing all trades outside the trading hours of the continuous market session, and in addition all trades that are within a minute after opening, or a minute before closing of the market. We also winsorize all variables at the 99% level to remove any effects of potential outliers, i.e. extreme volume days.

3.2.3 Trading Volume

The focus of our research is on trading volume, and more specifically, on opaque trading volume. Investigating executed volume instead of submitted volume is useful since it are executions that finally matter for market participants. Furthermore, the number of limit orders that is now unexecuted because of cancellations has grown dramatically over the last years (Hasbrouck and Saar, 2009).

[7] Some dark trading venues may have proprietary reporting arrangements, and therefore, may not be included in the sample.

Total *executed* trading volume contains four components: visible trading volume $VisV_{i,t}$, hidden order trading volume $HidV_{i,t}$, dark trading volume $DarkV_{i,t}$, and block trading volume $BlockV_{i,t}$. Subscript i refers to the stock and t refers to days since we aggregate our data at a daily frequency. These measures are the key variables in our analysis.

To develop a proxy for hidden order trading, we compute executed hidden order volume for stock i on lit venue l during day t, denoted $HidV_{l,i,t}$. We do this by matching data on executed trades with limit order book updates. When a market order executes against a limit order on lit venue l, this affects depth outstanding in the limit order book on l. If it is executed against a visible order, visible depth at that price is reduced by the same amount as the volume of the order. In contrast, if the market order executes against a (partially) hidden order, visible depth is not reduced by that same amount. In fact, visible depth may not be reduced at all due to a new peak replenishment of a reserve order. The excess of the order volume that was not executed against visible depth, is executed against (part of) a hidden order. This volume is categorized as hidden order volume on lit venue l: $HidV_{l,i,t}$.[8] Hidden order trading volume consolidated over all lit venues is then defined as $HidV_{i,t} = \sum_l HidV_{l,i,t}$.[9]. By analogy, the visible trading volume $VisV_{l,i,t}$ for stock i on day t that is executed on venue l is calculated as the trading volume that is not executed against hidden depth, and which thus can be matched to a limit order book update. Visible trading volume consolidated over all trading venues l is denoted by $VisV_{i,t}$.

Because we find a large dispersion in size for MiFID reported transactions we segment this volume (which stems from dark venues) into two categories: dark trading, denoted $DarkV_{i,t}$, and block trading, denoted $BlockV_{i,t}$. The first category comprises of transactions of 'regular size', more or less in line with the transaction size on lit venues (see Panel B of Table 3.1), while a second category consists of very large trades that are unlikely to find execution as a whole on lit venues.[10] We include in the first category all transactions that are smaller than 1% of the average daily euro volume on lit venues over the last 30 trading days. All other dark trades we consider to be 'block trades'. Although we can-

[8]Do note the difference between our measure of *executed* hidden order volume on lit venues and the volume of hidden orders *submitted* on lit venues. While the submitted hidden order volume on the lit venues can be substantial, volume that is executed against the hidden part of the order book is in the range of 4% to 6% of trading volume executed on lit venues, according to Panel C of Table 3.1. The difference stems from the fact that not all submitted hidden orders find execution, but also not all orders that were *submitted* hidden are *executed* when they are hidden. As the visible part of a reserve order is executed the peak size is replenished. Orders which were originally hidden may find themselves visibly executed if traders each time only trade against the visible part of the order book.

[9]Yao (2012) uses a similar measure of hidden order trading and shows that on average 19% of shares is executed against hidden depth on NASDAQ. Other studies also use a combination of order book data and transaction data to infer hidden order executions. Their focus, however, is on the detection of hidden orders by other market participants (see, e.g., De Winne and D'Hondt, 2007; Frey and Sandas, 2009; Pardo and Pascual, 2012)

[10]Depending on the size of the transaction relative to the Normal Market Size of stocks, large trades can also be eligible for delayed reporting.

not identify with certainty which trades are executed in dark pools, SIs or the OTC market, we believe that a classification based on size can make a meaningful distinction here. Regular sized dark transactions are more likely to be executed through automated trading venues such as dark pools, while big blocks of shares are likely executed through designated block trading mechanisms. Typically these block trades are negotiated bilaterally (over the phone or electronic networks) or matched in special block crossing networks, such as provided by Liquidnet. For a large part of the analysis we exclude block trading volume because it tends to be volatile and dominated by extreme observations. In addition, hidden orders on lit venues and dark trading venues are more likely to compete with each other for similar 'regular size' transactions, and not block size transactions.

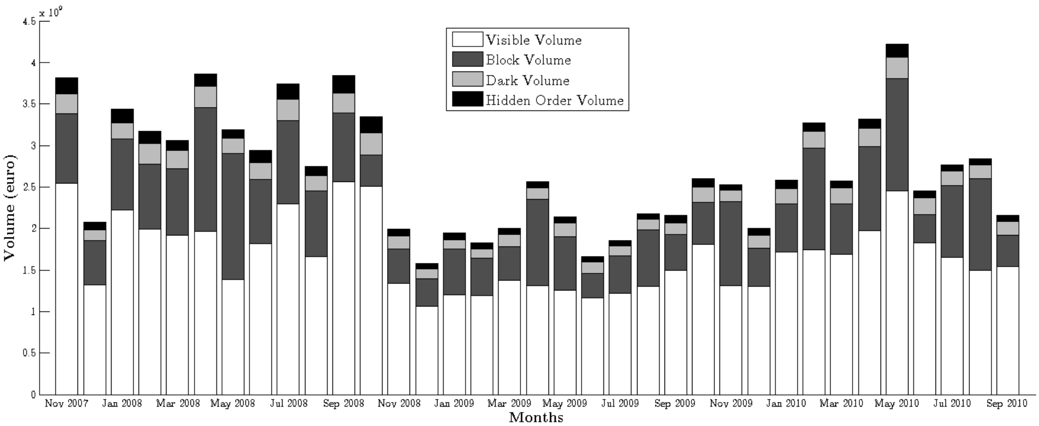
Figure 3.1 shows the monthly cross-sectional average and median levels of the four volume components, measured in euro, while Figure 3.2 present the daily cross-sectional average and median levels of the four volume components, measured relative to the total trading volume. We observe a considerable decline in total trading volume as of November 2008 due to the financial crisis. Visible volume is clearly the largest component of trading volume (about 64% on average), while block volume represents around 27% of volume. Dark volume and hidden order volume are relatively smaller, representing around 7% and 4% of trading volume on average, respectively. The division of volume in the four components remains relatively stable over the sample period. However, the volume share of hidden order trading diminishes slightly over the sample period, while the share of dark trading increases somewhat, but to a lesser extent. Overall, it is clear that both types of dark trading do not necessarily move together over time. This observation is at the core of our research questions on how both types of dark trading relate, under which conditions market participants favor the use of one type over another, and under which conditions they are reinforcing each other.

Table 3.1 Panel B further presents statistics on the daily euro volume, average trade size and number of trades for the four different components of trading volume. We compute each variable on a stock-day basis and then take the daily mean and median over all stocks. Total daily transaction volume is 129 million euro on average, but only 60 million euro for the median stock-day observation. The large standard deviations for volume measures (183 million euro for total volume), and the considerable difference between the 95th and 5th percentile (483 million euro for total volume) imply that there is considerable variation in volume across stocks and over time. We control for unobserved heterogeneity driving these differences by employing a standardization procedure to the data later on that takes into account stock-quarter fixed effects. Furthermore, because of the large scale differences in volume observations across stocks the standardization re-expresses volume and other variables in terms of standard deviations.

Figure 3.1: Monthly Volume Levels

This figure presents the monthly cross-sectional average and median values for the levels of the four components of trading volume that we define: hidden order volume (the volume of trades executed against hidden depth on lit trading venues), dark trading volume (volume of trades executed on completely dark trading venues, excluding large blocks), visible volume (the volume of trades executed against displayed depth on lit trading venues) and block volume (volume of block trades executed on completely dark trading venues).

(a) Average Monthly Volume



(b) Median Monthly Volume

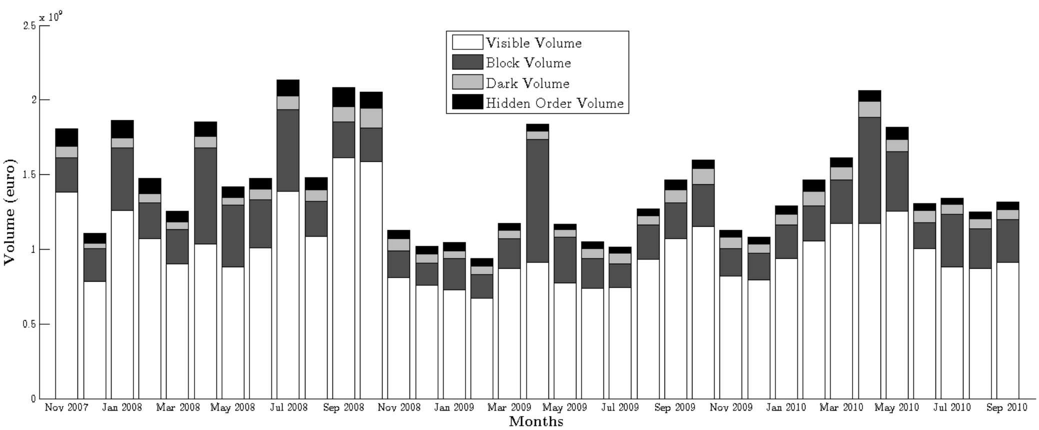
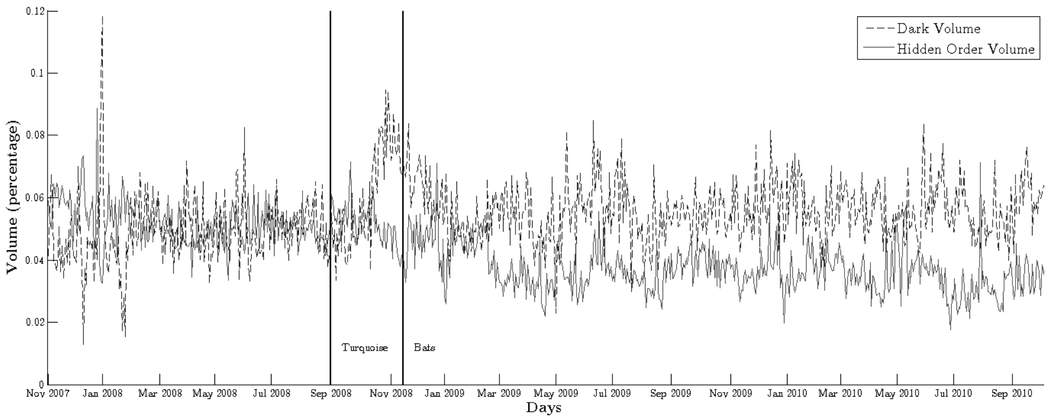


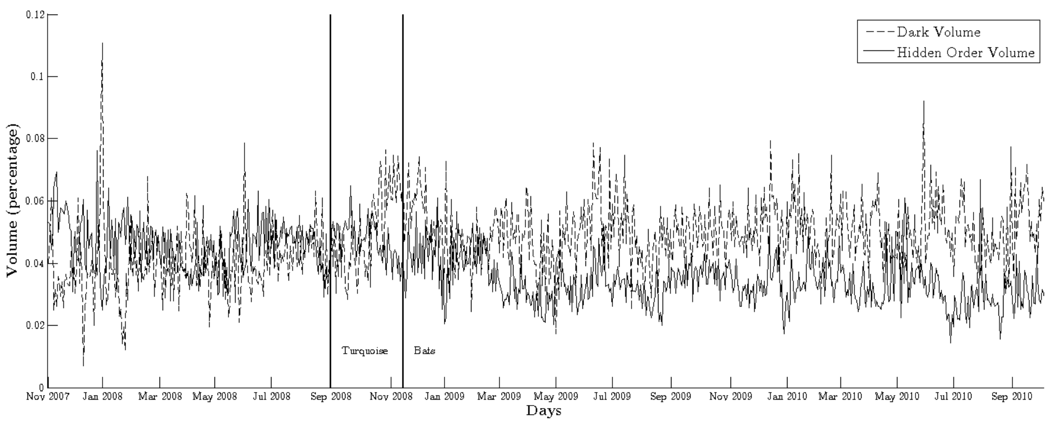
Figure 3.2: Daily Trading Volume Relative to Total Volume

This figure presents the daily cross-sectional average and median values for the relative levels of the four components of trading volume that we define. Panels (a) and (b) show the average and median hidden order volume (the volume of trades executed against hidden depth on lit trading venues) and dark trading volume (volume of trades executed on completely dark trading venues, excluding large blocks). Panels (c) and (d) show the visible volume (the volume of trades executed against displayed depth on lit trading venues) and block volume (volume of block trades executed on completely dark trading venues). The first vertical line depicts the entry of Turquoise in the Dutch stock market for most stocks in the sample, the second vertical line depicts the entry of Bats for most stocks.

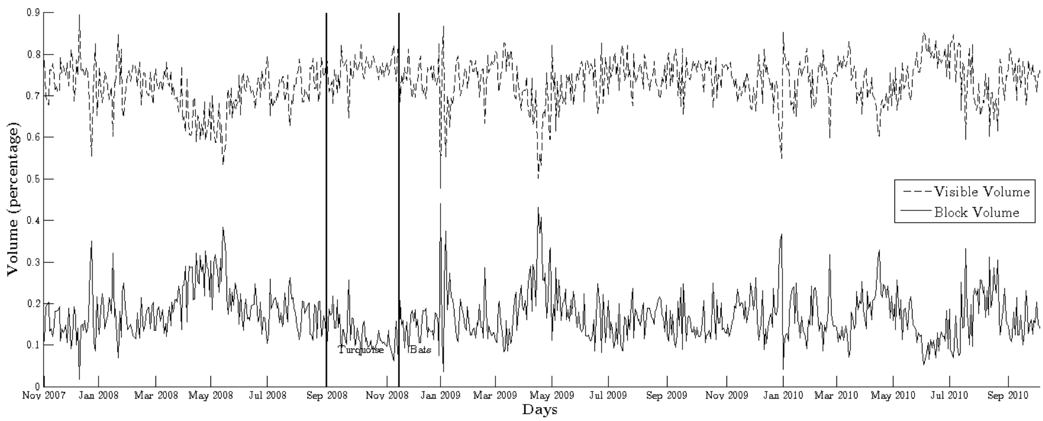
(a) Average Daily Relative Dark and Hidden Order Volume



(b) Median Daily Relative Dark and Hidden Order Volume



(c) Average Daily Relative Visible and Block Volume



(d) Median Daily Relative Visible and Block Volume

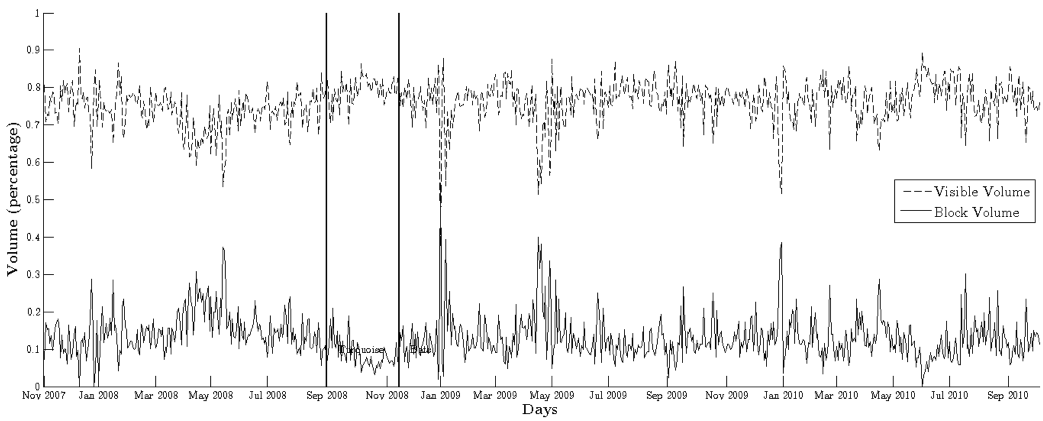


Table 3.1: Descriptive Statistics of Trading Volume

This table presents summary statistics based on daily observations for a number of volume measures. Panel A shows the number of stock-days for which there are observations in total, as well as for all the subsamples based on lit trading venues. MTF is the subsample for which at least one MTF is trading the stock; CHI is the subsample for which Chi-X is trading; TUR and BAT are similar subsamples for Turquoise and Bats respectively. Panel B shows statistics concerning the transactions at the consolidated market level: euro volume (in thousands), euro size, and number of trades. Panel C shows volume statistics across the different trading venues: lit venue euro volume, hidden order euro volume and the fraction of lit volume executed against hidden orders (in percentage points). In the first two blocks subsample volumes are expressed as a fraction of consolidated lit market volume (%, in percentage points). Subsamples are defined as follows: Main refers to the main trading venue, Alt refers to the consolidated alternative trading venues, CHI refers to Chi-X, TUR refers to Turquoise and BAT refers to Bats. Summary statistics in Panel C are based on observations for which at least one alternative venue trades the stock; for relative measures: when at least two alternative trading venues trade the stock, such that the total sums to 1.

Panel A: Number of Observations across Lit Venues

	Total	MTF	CHI	TUR	BATS
N	17,416	15,564	15,556	10,898	10,318

Panel B: Transactions at the Consolidated Level

	Mean	St. Dev.	P5	Med	P95
$Volume$					
Total	128,617	183,637	7,521	60,657	490,387
Lit	84,827	104,895	6,480	46,011	304,551
Visible	79,734	98,162	6,149	43,396	286,260
Hidden	5,093	7,749	182	2,410	19,009
Dark	8,614	15,105	102	2,783	38,795
Block	35,176	101,827	0	7,142	149,518
$Size$					
Lit	9,517	6,189	3,772	7,770	22,867
Visible	9,154	5,791	3,695	7,515	21,709
Hidden	12,300	10,155	3,678	9,332	32,233
Dark	57,833	86,934	5,067	25,920	225,038
Block	4,422,763	9,212,945	196,301	1,621,873	18,800,000
$NTrades$					
Lit	8,076	7,845	1,137	5,878	22,971
Visible	7,853	7,578	1,108	5,733	22,285
Hidden	358	453	36	243	1,029
Dark	175	228	7	98	606
Block	7	9	0	5	19

Panel C: Lit Volume across Venue Subsamples

	Mean	St. Dev.	P5	Med	P95
$Lit\ Volume$	89,458	104,754	7,257	50,501	309,267
%ENX	0.78	0.13	0.58	0.77	0.98
%MTF	0.22	0.13	0.02	0.23	0.42
%CHI	0.16	0.09	0.02	0.16	0.29
%TUR	0.04	0.04	-	0.03	0.12
%BAT	0.03	0.03	-	0.02	0.08
$Hidden\ Volume$	5,271	7,702	198	2,602	19,386
%ENX	0.82	0.18	0.43	0.88	0.99
%MTF	0.18	0.18	0.01	0.12	0.57
%CHI	0.10	0.11	0.00	0.07	0.30
%TUR	0.06	0.10	-	0.01	0.27
%BAT	0.02	0.04	-	0.01	0.10
$\%Hidden\ Volume$	0.05	0.03	0.02	0.05	0.11
ENX	0.06	0.03	0.02	0.05	0.11
MTF	0.04	0.06	0.01	0.03	0.13
CHI	0.04	0.06	0.00	0.02	0.12
TUR	0.07	0.11	-	0.03	0.26
BAT	0.06	0.13	-	0.02	0.21

Upon comparing trade size and the number of trades for block volume with the other volume components in Table 3.1 Panel B it becomes clear why it is distinctly different from other types of transactions. Block trades are limited to a mean (median) of 7 (5) transactions per day, while their average size is 78 times larger than that of other dark trades and much larger even than that of visible or hidden transactions. This provides us with a rational for removing these large blocks from the dark trading component.[11]

There are four lit trading venues in our data sample. Table 3.1 Panel C presents volume statistics spread out across the different venues. Euronext is the listing exchange and main trading venue by volume. Chi-X, Turquoise and Bats are three alternative venues (MTFs) of which Chi-X is the largest by volume and the only venue that trades stocks during the entire sample period. Because of these distinct differences we distinguish between the main trading venue ($Main$) on the one hand, and the consolidated alternative trading venues (Alt) on the other hand. Only a fraction of traders accesses the latter trading venues. We further distinguish these alternative venues into the largest competitor (Alt_L, i.e., Chi-X) and other trading venues (Alt_S, i.e. Turquoise and Bats), the smaller competitors which are not present during the full sample period, but only start trading during the last quarter of 2008.

3.3 Hidden Orders versus Dark Trading

3.3.1 Opaque Trading

The order exposure decision is crucial for most traders, especially so when they trade large sizes. The exposure decision pertains to how, where, when and to who trading intentions should be revealed to other market participants.[12] Order exposure has both benefits and costs associated with it (Harris, 1997). The primary benefit is that exposed orders can attract natural counterparties. Furthermore, in most lit trading venues opaque orders loose execution priority over visible orders, while execution probabilities in completely dark venues are generally also quite low. An increased execution probability is therefore also a major benefit of order visibility. However, there are also costs associated with order exposure or visibility because other traders (which do not represent a natural counter-

[11]Do note, however, that for visible and hidden order trading we are not comparing trade sizes of submitted limit orders. Large limit orders may be executed against several market orders. When part of a limit order is submitted as hidden, the limit order can be executed as hidden or visible. The trade sizes displayed here are that of marketable orders executing against the depth in the limit order book, either visible or hidden.

[12]In a broker-oriented trading structure where floor-based trading and electronic trading systems operate next to an upstairs market in which brokers arrange block trades, the exposure decision is related to that of broker selection. With the increased automation of financial market infrastructure, the relative importance of opaque orders submitted to electronic trading systems has grown. In fact, in automated venues opaque orders are substitutes for some services previously provided by floor brokers. As multiple automated venues are now competing for order flow, the order exposure decision today entails order selection as well as venue selection.

party) can react to the exposed trading intention. Bessembinder, Panayides, and Venkataraman (2009) show empirically that fully visible orders have higher execution costs compared to only partially visible orders.

Exposure costs of visible orders can arise for various reasons. First, traders that have trading intentions on the same side of the market could engage into more aggressive trading strategies as a reaction to the visible order. Buti and Rindi (2013) show that reserve orders, which reduce the visible part of a limit order, can help to avoid undercutting by aggressive traders. Second, exposure of trading intentions may lead other traders to engage into parasitic or predatory trading strategies that increase trading costs for the original trader (Harris, 1997; Brunnermeier and Pedersen, 2005). When predatory traders can easily detect large unfilled orders they can front-run these orders by taking liquidity from the market ahead of the original trader.[13] The third motive to reduce order visibility relates to the option value embedded in limit orders (Copeland and Galai, 1983). When visible orders provide committed trading options to other market participants and monitoring the market and adjusting order submissions is costly, these orders can be picked off by better informed traders (Foucault, 1999). Aitken, Berkman, and Mak (2001) argue that reducing exposure complicates strategies aimed at exploiting the free option value of limit orders because there is more uncertainty about the total order size. Finally, exposing trading intentions may cause prices to move in an unfavorable direction because traders assign an informational value to other traders' intentions (Harris, 1997). Moinas (2010) finds in a theory model that both informed and uninformed traders hide part of their limit orders to soften their informational impact. Boulatov and George (2013) show that informed traders compete more intensely as liquidity suppliers when they can hide their orders. In the experimental market designs of Gozluklu (2014) and Bloomfield, O'Hara, and Saar (2015) informed traders submit opaque orders in an effort to maintain their informational advantage for a longer period.

In a multi-venue setting, where lit and dark venues compete for order flow, the exposure decision is typically not explicitly modeled in the literature. Order choice in these models is between a costly market order on a traditional exchange, or a cheaper dark venue order. The former is more costly because it provides certainty of execution, whereas execution in a dark venue is more uncertain. Hendershott and Mendelson (2000), Degryse, Van Achter, and Wuyts (2009) and Daniëls, Dönges, and Heinemann (2013) model the lit venue (or traditional

[13]In doing so, they drive the market prices in an unfavorable direction and can make a profit later on by trading against the original trader at these worse prices. A front-running strategy that is specifically applicable to visible limit orders is quote-matching (Harris, 1996). Quote-matchers try to trade in front of large limit orders. Similar to other speculative strategies, they earn a profit when prices move in a favorable direction, while they remain protected from incurring heavy losses by the free options embedded in the limit order. Limit order traders are harmed because quote-matchers reduce their execution probabilities and at the same time try to extract the option value from their limit order.

exchange) as a dealer market in which traders pay the half-spread for a guaranteed execution of their order. In the models by Ye (2011) and Zhu (2014) buy and sell orders (including those of informed traders) are balanced and the heavier side incurs a market impact cost. On the lit venue, market impact translates into a price impact, while on the dark venue market impact results in non-execution (consistent with empirical evidence by Gresse (2006) and Naes and Odegaard (2006)). Although not explicitly modeled, the impact of exposure is implicit in the greater market impact of visible orders on lit venues. A similar reasoning is followed by Menkveld, Yueshen, and Zhu (2015) who put forward a pecking order theory of trading venues. They argue that investors trade off trading costs with their desire for immediacy, for different types of dark and lit trading venues. The most patient traders transact as much as possible in the cheapest (midquote matching) dark venues, while the most impatient traders choose for the immediacy offered by lit trading venues.

Hidden orders and dark trading venues have in common that they can be used to reduce exposure. Furthermore, for both opaque trading types reduced exposure also comes at the cost of lower execution probabilities. Both hidden order and dark trading can be considered passive strategies, i.e. they do not actively react to liquidity offered by other traders, and there is no certainty of execution. Instead, hidden orders and dark orders both provide opaque liquidity to other market participants. Uncertainty of execution is further raised because their opaque nature makes it more difficult to attract counterparties.

Besides similarities there are also important differences between both types of opaque trading. The main difference is that hidden orders interact with marketable orders on lit venues, while dark orders interact with similar dark orders. As such, hidden orders execute against more aggressive active orders. These are submitted by traders willing to pay for certainty of execution. Because hidden orders execute against marketable orders on lit venues they are also more easily detectable, and upon detection their presence is revealed to all traders monitoring the market (De Winne and D'Hondt, 2007; Frey and Sandas, 2009; Pardo and Pascual, 2012). Dark orders are relatively more opaque. To detect a dark order traders need to *ping* for dark liquidity on each of the dark venues separately. When execution in a dark venue is not continuous detecting dark orders is even more challenging. Furthermore, access to dark trading venues can be restricted, while most traders can use some form of hidden orders.

The remainder of this section investigates the interaction between hidden order trading and dark trading, controlling for market conditions. In particular, we ask the question whether hidden order trading and dark trading behave as complements or substitutes. Dark trading and hidden order trading can positively affect each other as both are driven by an underlying desire to trade opaquely. When traders with an opaque trading interest have access to both opaque trading tools they may decide to split their opaque orders over these alternatives.

When hidden order trading and dark trading are highly similar, we expect order submissions and executions using these opaque trading tools to be as well. As a result, both components of opaque trading volume would move together and we expect a positive coefficient. However, this does not necessarily imply a causal relation in the sense that traders are more inclined to trade in a dark venue when hidden volume on lit venues is increasing or vice versa. The underlying driver of opaque trading volume is the latent interest in opaque trading, which is in turn fueled by the general trading interest and the desire to reduce order exposure costs.

Contrary to the complementary behavior of the opaque trading components is the assertion that dark trading and hidden order trading are natural substitutes. Even though both types of opaque trading share similarities, under certain circumstances traders may favor one over the other and therefore do not evenly split orders over opaque trading alternatives. In addition, when an opaque order is executed in, say, a dark venue, traders could cancel their opaque orders on lit venues. Or they may resubmit their unexecuted orders from the dark venue to a lit venue as hidden orders. Under such circumstances, both types of opaque trading could indeed serve as substitutes for each other. When sufficient controls that could proxy for the latent general and opaque trading interest are added to the model, executed dark trading can negatively affect executed hidden order trading and vice versa.

3.3.2 Methodology

For each stock we examine the aggregated volume of trading on dark venues and the aggregate volume of hidden order trading that takes place on the various lit venues that trade in these stocks. To quantify the contemporaneous impact of two variables that are determined simultaneously one needs to employ a simultaneous equation model. The two endogenous variables of interest, dark trading activity and hidden order trading activity, are modeled as a function of each other and variables that are designed to capture prevailing order book and market conditions that potentially impact the volume measures. As these control variables themselves are potentially endogenous these are also instrumented.

To test whether dark trading and hidden order trading behave as complements or substitutes we estimate the following panel system of simultaneous equations:

$$\begin{cases} DarkV_{i,t} = \beta_{1,1} HidV_{i,t} + \beta_{1,2} VisV_{i,t} + \alpha_1 DarkV_{i'\neq i,t} + \boldsymbol{\gamma}_1' \mathbf{X}_{i,t} + \boldsymbol{\lambda}_1' \mathbf{Z}_{i,t} + \upsilon_{i,t} \\ HidV_{i,t} = \beta_{2,1} DarkV_{i,t} + \beta_{2,2} VisV_{i,t} + \alpha_2 HidV_{i'\neq i,t} + \boldsymbol{\gamma}_2' \mathbf{X}_{i,t} + \boldsymbol{\lambda}_2' \mathbf{Z}_{i,t} + \eta_{i,t} \\ VisV_{i,t} = \beta_{3,1} DarkV_{i,t} + \beta_{3,2} HidV_{i,t} + \alpha_3 VisV_{i'\neq i,t} + \boldsymbol{\gamma}_3' \mathbf{X}_{i,t} + \boldsymbol{\lambda}_3' \mathbf{Z}_{i,t} + \omega_{i,t} \end{cases} \tag{3.1}$$

An intercept is not included because all variables in the model are standardized by stock and quarter: we subtract the stock-specific mean of the given quarter of which day t is a part, and divide by the standard deviation. De-meaning the variables allows to control for unobserved heterogeneity in volume components across stocks by quarter. Because we are measuring our variables as deviations from their stock-quarter specific means, our estimation procedure in effect exploits the time series variation in volume components within each quarter. By dividing the variables by their standard deviation we normalize their units to standard deviations. This allows for an easy economic interpretation of results and more comparability across firms. A similar procedure is used by Buti, Rindi, and Werner (2011) and Hasbrouck and Saar (2013).

Our empirical proxies for the different components of trading volume are defined in subsection 3.2.3. We use the levels of volumes as opposed to volumes relative to the total executed volume because we are interested in the interplay between trading volume categories, which is quantified by the β coefficients. As mentioned before, a positive β indicates complementary behavior, while a negative β is an indication of substitution between the types of trading. If relative volumes are used a mechanical negative correlation is built in into the analysis because volume not executed through one of these trading alternatives is executed through one of the two other mechanisms.

As the components of trading volume are simultaneously determined we need instruments that are excluded from the other regressions in the system. We use as instruments their *market* levels, i.e. for each stock i and each day t we calculate the cross-sectional average of the volume components $DarkV_{j,t}$, $HidV_{j,t}$ and $VisV_{j,t}$ over all stocks $j \in 1,...,N$ excluding stock i. This results in the instruments $DarkV_{i' \neq i,t}$, $HidV_{i' \neq i,t}$ and $VisV_{i' \neq i,t}$. The positive correlation in different volume measures between stock i and the other stocks in the sample ('the market') on day t reflects a commonality in opaque and visible trading volumes. Intuitively, e.g., a correlated desire to trade in dark venues might be capturing a latent demand for the services that dark venues provide. In turn, this might be driven by institutional investors who make trading decisions for portfolios of stocks at the same time. Their routing and execution strategies are likely to be correlated across stocks. The inclusion of the exogenous 'market' volumes in each of these equations, that is excluded from the other equations, ensures that the system is identified. A similar approach is used by Buti, Rindi, and Werner (2011) to instrument dark trading, by Degryse, de Jong, and van Kervel (2015) to instrument dark trading and visible fragmentation, and by Hasbrouck and Saar (2013) to instrument low-latency trading. Following Hasbrouck and Saar (2013) we further exclude stocks that are in the same industry to reduce the potential of reversed causality even more.

The vectors $\mathbf{X}_{i,t}$ and $\mathbf{Z}_{i,t}$ contain endogenous and exogenous market and order book conditions that we identify as having a potential impact on opaque or visible trading activity. We discuss their effect into detail in Section 3.4. A similar approach is used to instrument the endogenous market and order book conditions by their 'market' versions.

As a robustness test we include block trading volume. Large blocks are transacted through designated mechanisms for which hidden order trading is probably a poor substitute. We modify Equation (3.1) to include $BlockV_{i,t}$, which is defined as the euro volume of block trades that is executed in dark venues for stock i on day t. These transactions comprise of the truly big blocks of shares that are usually negotiated or executed by a block crossing network.

$$\begin{cases} DarkV_{i,t} = & \beta_{1,1}HidV_{i,t} + \beta_{1,2}VisV_{i,t} + \beta_{2,3}BlockV_{i,t} + \alpha_1 DarkV_{i'\neq i,t} \\ & +\gamma_1'\mathbf{X}_{i,t} + \boldsymbol{\lambda}_1'\mathbf{Z}_{i,t} + \upsilon_{i,t} \\ HidV_{i,t} = & \beta_{2,1}DarkV_{i,t} + \beta_{2,2}VisV_{i,t} + \beta_{3,3}BlockV_{i,t} + \alpha_2 HidV_{i'\neq i,t} \\ & +\gamma_2'\mathbf{X}_{i,t} + \boldsymbol{\lambda}_2'\mathbf{Z}_{i,t} + \eta_{i,t} \\ VisV_{i,t} = & \beta_{3,1}DarkV_{i,t} + \beta_{3,2}HidV_{i,t} + \beta_{3,3}BlockV_{i,t} + \alpha_3 VisV_{i'\neq i,t} \\ & +\gamma_3'\mathbf{X}_{i,t} + \boldsymbol{\lambda}_3'\mathbf{Z}_{i,t} + \omega_{i,t} \\ BlockV_{i,t} = & \beta_{4,1}DarkV_{i,t} + \beta_{4,2}HidV_{i,t} + \beta_{4,3}VisV_{i,t} + \alpha_4 BlockV_{i'\neq i,t} \\ & +\gamma_4'\mathbf{X}_{i,t} + \boldsymbol{\lambda}_4'\mathbf{Z}_{i,t} + \vartheta_{i,t} \end{cases} \quad (3.2)$$

3.3.3 Results

Table 3.2 presents the estimation results of Equation (3.1). We estimate six specifications: a full specification under header (1), and five specifications that exclude one or several control variables. We follow this approach because the instrumentation procedure may decrease the power of our statistical tests. However, results are fairly robust across specifications. We find that hidden order trading and dark trading are negatively related to each other, despite both being driven by an unobservable opaque trading desire in the stock. Hidden order volume is significantly negatively affected by executed dark volume, while dark volume is also negatively affected by hidden order trading activity, but not significantly. This suggests that both types of opaque trading activity substitute for each other, rather than that they behave as complements. To understand this relation, recall that executed volume (or any subdivision or category of volume) has three determinants: trading desire, submission rate and execution rate. The inherent trading desire in a stock is likely to be determined by the same (exogenous) factors for both opaque and visible order trading volume (e.g., speculation, fundamentals or liquidity reasons). Next, the decision to reduce exposure to the market and thus to submit an opaque order type is driven by the same considerations, irrespective of the chosen type of opaque order. Because both types of opaque trading are strongly determined by the trading desire in general and the opaque trading desire in particular both volume measures tend to be correlated positively. However, we control for these trading interests by including the visible volume for the stock and the market levels of dark or hidden order volume (our instruments). Because the different types of opaque orders share the important similarity that they reduce exposure, but may further differ on other characteristics, hidden orders and dark orders are natural substitutes in a second stage decision, *after* deciding to trade opaquely. For instance, traders may choose one type of opaque trading over another based on perceived market conditions on the lit venues versus dark venues.

Another aspect that contributes to this result is the existence of a liquidity externality (Pagano, 1989). When opaque volume gravitates to dark venues the perceived liquidity on these venues increases and traders who prefer to trade opaquely become more inclined to route their orders there, further increasing dark trading activity. Alternatively, when traders perceive liquidity to be low in dark venues, because they observe lower volumes or worse than expected execution rates, they become more willing to choose an alternative method of trading opaquely, e.g. by re-routing their order to a lit venue as a hidden order. The effect of (perceived) execution probability thus reinforces any substitution effect in the submission phase. This effect is not entirely similar for dark venues and hidden orders, because in dark venues both sides of the trade constitute of traders with a desire to trade opaquely, which causes the liquidity externality on dark venues. For hidden orders the trader who wants to reduce his exposure is only on one side of the trade, while the other side constitutes of a more aggressive trader. Hence, the execution probability of hidden orders is not only determined by traders with a desire to trade opaquely.

The effect of dark volume on hidden order volume is economically smaller than the effect of hidden order volume on dark volume (-0.043 compared to -0.127 in the full specification), and insignificant. Dark volume can thus be a good substitute to hidden order volume, while substitution effects from hidden orders to dark venues are less likely. A potential explanation is that there remain to be important differences between both types of opaque trading that make substitution from dark venues to hidden orders less likely. The most important difference is that hidden orders are far less opaque than orders on dark venues, because the former are more easily detectable. When a marketable order trades against a hidden order on a lit venue it is immediately revealed to all traders that monitor the market that there is hidden liquidity available. As such, other traders often adjust their strategies by trading more aggressively against these orders (see, e.g., De Winne and D'Hondt, 2007; Frey and Sandas, 2009; Pardo and Pascual, 2012). Because dark venues do not have publicly displayed order books (by definition) traders that have an interest in detecting dark liquidity need to exert much more effort.

Hidden volume also negatively affects visible volume, as hidden and visible orders substitute for each other within the order books of lit venues. Hidden and visible limit orders share similarities because they are executed within the same

Table 3.2: Hidden Order versus Dark Trading: Simultaneous Equation System

This Table presents estimation results of the system of Equations (3.1) for different specifications. The endogenous dependent variables are $DarkV$ the euro volume of non-block trades executed across all dark venues relative to the total consolidated volume; $HidV$, the euro volume of hidden orders executed across all lit venues; and $VisV$, the euro volume of visible orders executed across all lit venues. Independent variables in the models are: $QSpread_{Lit}$, the time-averaged consolidated market quoted bid-ask spread (relative to the midquote); $VisDepth_{Lit}$, the time-averaged consolidated lit market depth quoted within 50 basis points of the midquote, based on Degryse, de Jong, and van Kervel (2015); $Volat$, the daily standard deviation of five-minute midquote returns; AT_{Lit}, a proxy for algorithmic trading based on Hendershott, Jones, and Menkveld (2011), and SOR, an estimate of the fraction of traders employing smart order routers, based on van Kervel (2015). $V_{i' \neq i}$ is the instrument for the different volume measures, either $DarkV_{i' \neq i,t}$, $HidkV_{i' \neq i,t}$ or $VisV_{i' \neq i,t}$. All equations are estimated by 2SLS on a panel of daily observations for 27 Dutch large cap stocks which spans 738 trading days, from November 2007 until September 2010. Variables are standardized by stock-quarter by subtracting the mean and dividing by the standard deviation. Endogenous independent variables $QSpread_{Lit}$, $VisDepth_{Lit}$ and $Volat$ are instrumented by using their 'market version', i.e. for each stock i and day t we take the cross-sectional mean of the variable excluding stock i. AT_{Lit} and SOR are assumed to be exogenous. Standard errors are Newey-West for panel data, t-statistics are shown in parentheses, and ***, ***, * indicates significance at the 1%, 5% and 10% level respectively.

	(1)			(2)			(3)		
	$DarkV$	$HidV$	$VisV$	$DarkV$	$HidV$	$VisV$	$DarkV$	$HidV$	$VisV$
$DarkV$		-0.04	-0.04		-0.04	-0.02		-0.04	-0.04
		(-1.22)	(-1.03)		(-0.96)	(-0.41)		(-1.23)	(-0.99)
$HidV$	-0.127**		-0.212***	-0.05		-0.03	-0.140***		-0.246***
	(-2.35)	(0.00)	(-3.26)	(-0.95)		(-0.61)	(-2.68)		(-3.80)
$VisV$	0.150***	0.175***		0.177***	0.229***		0.154***	0.174***	
	(3.55)	(4.84)	(0.00)	(4.25)	(6.55)		(3.70)	(4.85)	
$QSpread_{Lit}$	0.022	-0.004	0.053**	0.007	-0.025	0.022			
	(1.09)	(-0.22)	(2.55)	(0.34)	(-1.42)	(1.10)			
$VisDepth_{Lit}$	0.038*	0.018	0.068***	0.016	-0.023	0.017	0.029*	0.019	0.047***
	(1.89)	(0.98)	(3.35)	(0.81)	(-1.24)	(0.80)	(1.68)	(1.21)	(2.57)
$Volat$	0.129***	0.176***	0.282***	0.016	0.010	0.026	0.140***	0.175***	0.313***
	(5.72)	(8.63)	(11.72)	(0.81)	(0.55)	(1.30)	(6.56)	(8.69)	(14.21)
AT_{Lit}	-0.256***	-0.377***	-0.583***				-0.257***	-0.378***	-0.589***
	(-14.29)	(-20.08)	(-16.93)				(-14.28)	(-20.87)	(-16.93)
SOR	0.031***	-0.033***	0.037***				0.030***	-0.033***	0.035***
	(3.62)	(-4.88)	(4.55)				(3.52)	(-4.87)	(4.28)
$V_{i' \neq i}$	0.596***	0.542***	0.870***	0.604***	0.615***	0.919***	0.597***	0.543***	0.878***
	(22.85)	(14.86)	(22.32)	(23.27)	(16.26)	(22.48)	(22.82)	(15.19)	(22.52)
Obs	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416
R^2	0.185	0.465	0.404	0.172	0.347	0.238	0.183	0.465	0.395

Table 3.2 continued.

	(4)			(5)			(6)		
	$DarkV$	$HidV$	$VisV$	$DarkV$	$HidV$	$VisV$	$DarkV$	$HidV$	$VisV$
$DarkV$		-0.040	-0.03		-0.030	-0.024		-0.031	-0.022
		(-1.16)	(-0.81)		(-0.87)	(-0.54)		(-0.90)	(-0.47)
$HidV$	-0.151***		-0.263***	-0.125**		-0.243***	-0.166***		-0.381***
	(-2.90)		(-3.94)	(-2.32)		(-3.04)	(-3.15)		(-4.09)
$VisV$	0.171***	0.182***		0.222***	0.272***		0.262***	0.296***	
	(4.32)	(5.31)		(5.48)	(8.85)		(6.83)	(10.34)	
$QSpread_{Lit}$	0.005	-0.011	0.025	0.075***	0.069***	0.200***			
	(0.29)	(-0.75)	(1.27)	(3.99)	(3.95)	(8.85)			
$VisDepth_{Lit}$				0.012	-0.018	0.012			
				(0.60)	(-1.09)	(0.54)			
$Volat$	0.120***	0.170***	0.276***						
	(5.41)	(8.87)	(11.13)						
AT_{Lit}	-0.258***	-0.373***	-0.605***	-0.236***	-0.348***	-0.631***	-0.223***	-0.322***	-0.658***
	(-14.31)	(-20.97)	(-17.04)	(-14.28)	(-19.90)	(-14.54)	(-14.31)	(-22.16)	(-13.52)
SOR	0.031***	-0.033***	0.039***	0.031***	-0.033***	0.044***	0.024***	-0.040***	0.028***
	(3.64)	(-4.85)	(4.59)	(3.67)	(-4.89)	(4.69)	(2.92)	(-6.18)	(2.88)
$V_{i'\neq i}$	0.600***	0.536***	0.903***	0.602***	0.544***	1.028***	0.604***	0.530***	1.143***
	(22.98)	(15.30)	(22.51)	(23.08)	(14.47)	(18.84)	(23.20)	(14.71)	(18.14)
Obs	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416
R^2	0.183	0.471	0.368	0.189	0.480	0.238	0.190	0.495	0.152

trading systems against the same marketable order flow. In addition, reserve orders on lit venues may execute partly hidden and partly visible, as the visible peak size gets replenished during the trading process.

We estimate equation (3.2), which includes block volume as a fourth category of volume, using the same procedures as before. The main results are presented in Table 3.3 and are robust: our coefficients remain of a similar sign, economic magnitude and statistical significance. We also find that visible volume can be a substitute for block volume as visible volume is negatively affected by block volume. For instance, when there is insufficient liquidity in block trading mechanisms, block traders are forced to execute their trades in other trading venues, e.g., by splitting their order in the lit market. Dark trading and block trading are complementary because there is some overlap in trading venues between non-block dark trades and block trades. Therefore, both non-block dark trading and block trading are higher when there is an increased interest in trading away from regulated and transparent trading venues.

Lastly, it should be noted that there are alternative explanations possible for the substitution result. This is because we look at executed volumes, which have multiple drivers, of which order submission is only one component. Interpretation depends upon the assumptions that are made about which traders are more likely to switch between certain order types. A potential alternative explanation is that some traders substitute between orders on dark venues and market orders on lit venues. An increased use of the latter type of orders increases the execution probability of hidden orders and therefore may also increase executed volumes of hidden orders. However, we believe that substitution between dark venue orders and hidden orders is more likely than substitution between dark venue orders and market orders. First and foremost, dark venue orders and hidden orders are both order types that allow traders to trade opaquely and not reveal their trading intention. Secondly, both hidden orders and dark venue orders are relatively passive in nature, are less costly in terms of the bid-ask spread and price impact, and have an uncertain execution. Market orders, by contrast, are costly orders with certain execution. This is in line with Buti, Rindi, and Werner (2015) who show that, when a dark pool is added next to a limit order book, it are predominantly limit order traders who migrate to the dark pool.

3.4 What Drives Opaque Trading?

3.4.1 Market and Order Book Variables

We now turn to the question of how market conditions affact opaque trading activity. Most of the market and order book variables we consider have a consolidated total market version (i.e., grouped over all trading venues, lit and dark), denoted by the subscript Tot, a consolidated lit market version (i.e., grouped over all lit trading venues), denoted by the subscript Lit, and a lit venue-specific ver-

Table 3.3: Hidden Order versus Dark Trading: Simultaneous Equation System (with Block Volume)

This Table presents estimation results of the system of Equations (3.2) for different specifications. The endogenous dependent variables are $DarkV$ the euro volume of non-block trades executed across all dark venues relative to the total consolidated volume; $HidV$, the euro volume of hidden orders executed across all lit venues; $VisV$, the euro volume of visible orders executed across all lit venues; and $BlockV$, the euro volume of block trades executed across all dark venues. Independent variables in the models are: $QSpread_{Lit}$, the time-averaged consolidated market quoted bid-ask spread (relative to the midquote); $VisDepth_{Lit}$, the time-averaged consolidated lit market depth quoted within 50 basis points of the midquote, based on Degryse, de Jong, and van Kervel (2015); $Volat$, the daily standard deviation of five-minute midquote returns; AT_{Lit}, a proxy for algorithmic trading based on Hendershott, Jones, and Menkveld (2011), and SOR, an estimate of the fraction of traders employing smart order routers, based on van Kervel (2015). $V_{i'\neq i}$ is the instrument for the different volume measures, either $DarkV_{i'\neq i,t}$, $HidkV_{i'\neq i,t}$, $VisV_{i'\neq i,t}$ or $BlockV_{i'\neq i,t}$.

All equations are estimated by 2SLS on a panel of daily observations for 27 Dutch large cap stocks which spans 738 trading days, from November 2007 until September 2010. Variables are standardized by stock-quarter by subtracting the mean and dividing by the standard deviation. Endogenous independent variables $QSpread_{Lit}$, $VisDepth_{Lit}$ and $Volat$ are instrumented by using their 'market version', i.e. for each stock i and day t we take the cross-sectional mean of the variable excluding stock i. AT_{Lit} and SOR are assumed to be exogenous. Standard errors are Newey-West for panel data, t-statistics are shown in parentheses, and ***, ***, * indicates significance at the 1%, 5% and 10% level respectively.

	(1)				(2)				(3)			
	$DarkV$	$HidV$	$VisV$	$BlockV$	$DarkV$	$HidV$	$VisV$	$BlockV$	$DarkV$	$HidV$	$VisV$	$BlockV$
$DarkV$		-0.031	-0.016	0.088*		-0.036	-0.011	0.090*		-0.031	-0.012	0.088*
		(-0.82)	(-0.38)	(1.86)		(-0.89)	(-0.25)	(1.89)		(-0.83)	(-0.30)	(1.86)
$HidV$	-0.131**		-0.204***	-0.049	-0.057		-0.030	0.004	-0.144***		-0.235***	-0.047
	(-2.43)	(0.00)	(-3.15)	(-0.73)	(-1.12)		(-0.57)	(0.07)	(-2.77)		(-3.67)	(-0.72)
$VisV$	0.147***	0.176***		0.025	0.171***	0.229***		0.039	0.151***	0.176***		0.024
	(3.48)	(4.89)	(0.00)	(0.50)	(4.13)	(6.56)		(0.78)	(3.63)	(4.89)		(0.49)
$BlockV$	0.051	-0.038	-0.077**		0.076*	-0.001	-0.019		0.049	-0.038	-0.083**	
	(1.20)	(-1.10)	(-1.95)	(0.00)	(1.82)	(-0.03)	(-0.43)		(1.15)	(-1.09)	(-2.09)	
$QSpread_{Lit}$	0.023	-0.005	0.051**	-0.003	0.009	-0.025	0.021	-0.012				
	(1.15)	(-0.28)	(2.44)	(-0.14)	(0.46)	(-1.42)	(1.07)	(-0.57)				
$VisDepth_{Lit}$	0.033*	0.021	0.074***	0.049**	0.010	-0.022	0.018	0.032	0.024	0.022	0.054***	0.051***
	(1.65)	(1.13)	(3.60)	(2.23)	(0.52)	(-1.23)	(0.86)	(1.43)	(1.38)	(1.39)	(2.92)	(2.60)
$Volat$	0.129***	0.175***	0.279***	0.053**	0.021	0.010	0.025	-0.024	0.141***	0.173***	0.308***	0.052**
	(5.75)	(8.57)	(11.61)	(2.23)	(1.04)	(0.54)	(1.22)	(-1.14)	(6.64)	(8.61)	(14.00)	(2.27)
AT_{Lit}	-0.249***	-0.381***	-0.587***	-0.177***					-0.250***	-0.382***	-0.594***	-0.177***
	(-12.95)	(-19.57)	(-16.97)	(-9.03)					(-12.94)	(-20.26)	(-17.00)	(-9.03)
SOR	0.031***	-0.034***	0.036***	-0.005					0.030***	-0.034***	0.034***	-0.005
	(3.64)	(-4.96)	(4.43)	(-0.56)					(3.54)	(-4.95)	(4.17)	(-0.55)
$V_{i'\neq i}$	0.587***	0.544***	0.871***	0.578***	0.578***	0.615***	0.919***	0.588***	0.587***	0.546***	0.879***	0.578***
	(21.39)	(14.76)	(22.31)	(18.47)	(21.80)	(16.14)	(22.46)	(18.55)	(21.35)	(15.05)	(22.53)	(18.46)
Obs	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416
R^2	0.192	0.461	0.400	0.103	0.184	0.347	0.236	0.089	0.191	0.460	0.390	0.103

Table 3.3 continued.

	(4)				(5)				(6)			
	$DarkV$	$HidV$	$VisV$	$BlockV$	$DarkV$	$HidV$	$VisV$	$BlockV$	$DarkV$	$HidV$	$VisV$	$BlockV$
$DarkV$		-0.029	-0.012	0.092**		-0.009	0.022	0.094**		0.005	0.066	0.096**
		(-0.78)	(-0.28)	(1.95)		(-0.26)	(0.47)	(2.01)		(0.14)	(1.33)	(2.07)
$HidV$	-0.152***		-0.260***	-0.081	-0.128**		-0.227***	-0.046	-0.166***		-0.352***	-0.072
	(-2.92)		(-3.91)	(-1.23)	(-2.37)		(-2.87)	(-0.69)	(-3.15)		(-3.85)	(-1.10)
$VisV$	0.165***	0.185***		0.052	0.221***	0.273***		0.053	0.262***	0.297***		0.070
	(4.13)	(5.40)		(1.07)	(5.43)	(8.86)		(1.06)	(6.80)	(10.24)		(1.41)
$BlockV$	0.058	-0.033	-0.063		0.032	-0.065*	-0.141***		-0.007	-0.109***	-0.264***	
	(1.42)	(-0.97)	(-1.56)		(0.74)	(-1.86)	(-3.05)		(-0.18)	(-3.19)	(-5.29)	
$QSpread_{Lit}$	0.009	-0.013	0.021	-0.024	0.076***	0.066***	0.193***	0.019				
	(0.50)	(-0.89)	(1.06)	(-1.36)	(4.06)	(3.79)	(8.49)	(0.94)				
$VisDepth_{Lit}$					0.009	-0.013	0.025	0.039*				
					(0.45)	(-0.73)	(1.03)	(1.81)				
$Volat$	0.122***	0.168***	0.273***	0.043*								
	(5.51)	(8.77)	(11.02)	(1.83)								
AT_{Lit}	-0.250***	-0.376***	-0.611***	-0.179***	-0.232***	-0.354***	-0.638***	-0.167***	-0.224***	-0.335***	-0.675***	-0.171***
	(-12.93)	(-20.61)	(-17.04)	(-9.08)	(-12.81)	(-19.24)	(-14.58)	(-9.37)	(-12.57)	(-20.99)	(-13.57)	(-10.12)
SOR	0.031***	-0.034***	0.038***	-0.005	0.031***	-0.034***	0.043***	-0.005	0.024***	-0.041***	0.027***	-0.005
	(3.67)	(-4.92)	(4.51)	(-0.54)	(3.69)	(-5.02)	(4.49)	(-0.55)	(2.93)	(-6.19)	(2.74)	(-0.52)
$V_{i' \neq i}$	0.588***	0.537***	0.906***	0.585***	0.596***	0.548***	1.027***	0.573***	0.605***	0.536***	1.136***	0.580***
	(21.40)	(15.24)	(22.50)	(18.74)	(21.58)	(14.31)	(18.79)	(18.54)	(21.52)	(14.58)	(17.99)	(19.30)
Obs	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416
R^2	0.192	0.469	0.363	0.103	0.194	0.471	0.223	0.105	0.189	0.476	0.091	0.105

sion which is denoted by the subscript l. We distinguish between the main venue ($Main$), the consolidated alternative venues (Alt), and the largest and smallest alternative venues (Alt_L and Alt_S respectively). Table 3.4 provides descriptive statistics on a selection of the market and order book variables. The remainder of this subsection further discusses the potential impact of these variables on the various types of opaque trading activity, and how they are empirically proxied.

Volume. We include volume as a market variable and potential determinant of dark and hidden order trading behavior because it is a proxy for trading interest. Hidden orders are more likely to be executed when there is a heavier trading interest. When the market is active and trading interest is higher, impatient traders are more likely to trade beyond visible depth in the limit order book to execute their marketable orders. More trading volume thus increases the execution probability of hidden orders. Similarly, dark orders also have a higher execution probability when trading interest in dark venues is higher. However, when the larger trading interest is driven by an increased demand for immediacy, traders may favor lit venues over dark venues, reducing the market share of dark venues. Volume is either measured as total volume $Volume_{Tot,i,t}$, lit market volume $Volume_{Lit,i,t}$ or venue-specific volume $Volume_{l,i,t}$.

Volatility. Volatility increases the option value of limit orders, which provides incentives to hide the full order size. Tuttle (2006) shows empirically that hidden orders are used more for stocks that exhibit a higher volatility and idiosyncratic risk. But limit orders also tend to be less aggressively priced when volatility is larger because picking-off risk is greater (Foucault, 1999). This increases trading costs for marketable orders, which in turn could reduce executed hidden order volume at the cost of visible order volume. For dark venues, when volatility proxies for adverse selection, Zhu (2014) predicts that dark pool market share decreases when volatility increases. Volatility also increases the demand for immediacy and decreases execution probabilities in dark venues.

Volatility $Volat_{i,t}$ is measured only on the market level as the standard deviation of five-minute midquote returns using the midquote of the consolidated limit order book. Table 3.4 shows that our stock-specific daily volatility estimate $Volat_{i,t}$ is 23.45 basis points on average, with 9.01 basis points on low-volatility days in the 5th percentile and 50.42 basis points on high-volatility days in the 95th percentile.

Spread. Using hidden orders as tools to protect against predatory or competitive behavior of other traders is only useful when the bid-ask spread is sufficiently large. Therefore, hidden order submissions are expected to be more likely when the spread is larger. However, with regard to execution of hidden orders, a larger bid-ask spread also increases trading costs for marketable orders. In turn this reduces the execution probability of hidden depth more than that of visible depth. For dark venues, Buti, Rindi, and Werner (2015) predict that due to competition between liquidity providers a narrower spread induces traders

Table 3.4: Descriptive Statistics of Order Book and Market Variables

This Table presents descriptive statistics on market and order book variables. The first block shows visible depth $VisDepth$ measured in euro and the fraction of visible depth (%, in percentage points) that is quoted on selected subsamples of lit venues. Visible depth is defined as the consolidated lit market visible depth offered in an interval of 50 basis points around the midquote, based on Degryse, de Jong, and van Kervel (2015). The second block shows the consolidated quoted spread $QSpread$, relative to the midquote (expressed in basis points) across all trading venues, and that of selected subsamples of lit venues. Both visible depth and quoted spread are based on daily time-weighted averages derived from minute-by-minute order book snapshots. The third block shows the number of $Messages$ transmitted to the consolidated market, and the fraction of those messages that (%, in percentage points) is transmitted to selected subsamples of lit venues. The fourth block shows statistics for algorithmic trading AT at the consolidated lit market level, and AT on selected subsamples of lit venues. The AT measure is based on Hendershott, Jones, and Menkveld (2011). The next-to-last row presents statistics on the fraction of traders using SOR based on the γ_2 proxy of van Kervel (2015), while the last row presents statistics on volatility at the consolidated lit market level (expressed in basis points). Subsamples are defined as follows: Main refers to the main trading venue, Alt refers to the consolidated alternative venues, Alt_L refers to the largest alternative trading venue, Alt_S refers to the other alternative trading venues. Summary statistics in Panel D are based on observations for which at least one alternative venue trades the stock; for relative measures: when at least two alternative venues trade the stock, such that the total sums to 1.

	Mean	St. Dev.	P5	Med	P95
Visible Depth	409,022	439,459	31,311	273,526	1,270,413
%Main	51.46	8.78	36.83	51.94	65.69
%Alt	48.54	8.78	34.31	48.06	63.17
%Alt_L	26.35	6.41	14.95	26.91	35.47
%Alt_S	22.20	8.67	11.08	20.43	38.96
Quoted Spread	7.72	5.15	2.97	6.66	16.70
Main	10.38	6.13	4.11	8.95	21.89
Alt	12.96	18.94	3.88	9.03	34.16
Alt_L	16.34	24.67	4.34	10.29	47.63
Alt_S	17.65	17.69	6.70	12.88	41.45
Messages	378,313	422,321	31,268	236,088	1,188,858
%Main	31.67	10.97	12.54	32.82	48.48
%Alt	68.33	10.97	51.52	67.18	87.46
%Alt_L	30.32	9.35	16.60	30.00	46.22
%Alt_S	38.01	12.76	21.46	35.85	62.81
AT	-4.52	6.61	-19.37	-1.83	-0.41
Main	-8.95	11.14	-34.22	-4.45	-1.01
Alt	-0.96	0.93	-2.81	-0.69	-0.10
Alt_L	-1.21	0.98	-3.11	-0.94	-0.18
Alt_S	-0.44	0.48	-1.29	-0.30	-0.03
SOR	7.43	4.88	0.85	7.03	15.69
Volatility	23.45	15.57	9.01	19.60	50.42

to migrate away from lit venues. A narrower spread makes midquote execution (which is typical in many dark venues) relatively more attractive for patient traders. Alternatively, when primarily traders who would otherwise demand liquidity on lit venues substitute the latter for dark venues, dark trading should be increasing in the size of the spread (Hendershott and Mendelson, 2000; Degryse, Van Achter, and Wuyts, 2009).

$QSpread_{i,t,l}$ denotes the quoted spread between the best bid and ask quote for venue l; $QSpread_{i,t,Lit}$ is the consolidated lit market quoted bid-ask spread. The bid-ask spread is measured relative to the midquote. It is recorded from the order book based on one-minute snapshots and then time-weight these observations to have a daily measure. The consolidated quoted spread $QSpread_{i,t,Lit}$ is on average (median) 7.72 (6.66) basis points of the midquote, as can be judged from Table 3.4. $QSpread_{i,t,l}$ is the tightest on the main trading venue (10.38 basis points on average) and the widest on the smaller alternative trading venues (17.65 basis points on average).

Visible Depth. Because of the lower execution priority of hidden depth over visible depth, hidden depth is only executed insofar visible depth is depleted. Hence, when visible depth is larger, the volume share that is executed against hidden orders is expected to be lower. The lower execution probability further reduces incentives to submit hidden orders. Both De Winne and D'Hondt (2007) and Bessembinder, Panayides, and Venkataraman (2009) find that visible depth on the same side of the market reduces the probability of submitting a hidden order. In a market where multiple trading venues are competing and offering options to hide the full order size, more visible depth on one venue should induce large patient traders to submit their hidden orders on another venue. These traders are effectively crowded out on the most liquid venue and substitute it for a venue where less visible depth is offered. For dark venues the prediction from Buti, Rindi, and Werner (2015) with regard to the effect of depth is similar to that for the quoted spread: as more liquidity providers compete for execution in the lit market it becomes deeper, and some traders are crowded out to dark venues. On the other hand, deeper lit venues may attract order flow away from dark markets if primarily liquidity demanders substitute between dark and lit trading venues.

We use the $Depth\ (X)$ measure from Degryse, de Jong, and van Kervel (2015) as an empirical proxy for visible depth. It is constructed as follows:

$$\begin{aligned}
Depth^{ask}(X) &= \sum_{j=1}^{J} p_j^{ask} q_j^{ask} \mathbb{1}\left(p_j^{ask} < m(1+X)\right) \\
Depth^{bid}(X) &= \sum_{j=1}^{J} p_j^{bid} q_j^{bid} \mathbb{1}\left(p_j^{bid} > m(1-X)\right) \\
Depth(X) &= Depth^{ask}(X) + Depth^{bid}(X)
\end{aligned}$$

Where j denotes a level on the pricing grid of a venue and m the midquote of the consolidated order book across all lit venues. By measuring depth relative to the consolidated midquote m we take into account that only visible depth that is relatively close to the best prices of the consolidated market is relevant and competitive. We choose $X = 50$ basis points. For each venue l we denote *own* venue visible depth as $VisDepth_{i,t,l}$ and aggregated visible depth across the *other* lit venues $l' \neq l$ as $VisDepth_{i,t,l' \neq l}$. $VisDepth_{i,t,Lit}$ is the consolidated lit market visible depth. Similar to the quoted spread, visible depth is calculated from the limit order book based on one-minute snapshots and then time-weighted throughout the day to have a daily measure of depth.

In Table 3.4 we find that $VisDepth_{i,t,Lit}$ is on average 409,022 euro, with a median of 273,526 euro. The large range between the 5th and 95th percentile that was found for volume measures in Table 3.1, is also found for depth, with a 1,240,000 euro difference. On average a little bit over half of visible depth within 50 basis points is quoted by the main trading venue, the remainder by alternative trading venues. For the alternative trading venues more than half of depth is quoted by Chi-X, the largest competitor.

Algorithmic Trading. An important feature of today's markets is the use of algorithms by several classes of traders. For instance, brokers and buy-side traders use algorithms to optimally split and time their trades. An important class of algorithmic trading (AT) is high-frequency trading (HFT).[14] The presence of algorithmic traders (and the strategies they employ) could impact the use and execution probability of opaque orders. Since ATs are highly competitive when they provide liquidity hidden orders are not suitable order types because of their relatively low execution probability. Because ATs compete fiercly with each other and other liquidity providers for order execution, hidden order executions are likely to suffer from the presence of ATs. Furthermore, some AT strategies may be designed to front-run large traders or exploit their orders in some other way. If large traders fear the presence of ATs they may refrain from trading opaquely and seek execution of their orders elsewhere (i.e., using block trading mechanisms). This behavior reduces dark and hidden order trading. In addition, algorithms can also be used as a substitute to opaque trading by large traders to split up their order, which further reduces opaque trading.

Our proxy for algorithmic trading AT is related to the message-to-volume ratio (the number of messages that needs to be transmitted to transact one euro of volume). We use the measure from Hendershott, Jones, and Menkveld (2011) to proxy for AT on each lit venue l:

$$AT_{i,t,l} = -\frac{\frac{LitV_{i,t,l}}{100}}{Messages_{i,t,l}}$$

Where $LitV_{i,t,l}$ is the euro volume executed on the lit venue and $Messages_{i,t,l}$ is the number of electronic messages or updates in the limit order book until the tenth best price on either side of the market (which consists of order submis-

[14]High-frequency traders are proprietary traders who rely on algorithms to make their trading decisions and compete primarily on speed. They use their speed advantage for a diversity of trading strategies. Market making is one of the strategies for which HFT technology seems to be advantageous, as high-frequency traders are now responsible for a large part of liquidity provision (see, e.g., Menkveld, 2013; Hagströmer and Nordén, 2013; Brogaard, Hendershott, and Riordan, 2014).

sions, cancellations and modifications). AT is increasing in algorithmic trading. $AT_{i,t,Lit}$ is the same measure applied to the consolidated lit market.

On an average day, for the average stock, 378,313 messages are sent to lit trading venues, but this goes up to 1,188,858 messages for the 95th percentile of stock-days, as shown in Table 3.4. Less than a third is disseminated to the main exchange on average days, but again there is considerable variation with a P5-P95 range from 12.54 percent to 48.48 percent of total messages. A similar amount of messages is sent to the largest alternative venue (Chi-X) each day, with up to 38 percent of the messages sent to the smaller alternative venues. Because volumes are much lower on alternative trading venues, while the amount of messages sent is similar or even higher, our algorithmic trading proxy indicates that there is more algorithmic trading on these trading venues. The mean (median) of $AT_{i,t,l}$ is -8.95 (-4.45) for the main venue and -0.96 (-0.69) for the alternative trading venues.

Smart Order Routers. A fragmented market is characterized by how well its different trading venues are interconnected. In a European context in which there is no National Market System or trade-through prohibition this heavily depends on the use of Smart Order Routers (SOR) by traders and brokers. These are algorithms specifically designed to tap into the liquidity of multiple trading venues. The more traders use SOR, the more trading venues are competing with each other. The use of SOR can turn a fragmented market into a virtually consolidated market. Typically, a trader using SOR will first deplete depth at the trading venue that provides the best price, and then continue to the trading venue that offers the same price (but smaller depth), until his order is filled or all depth at the best price is depleted. If the order is not entirely filled it is send to the venue quoting the second best price and so on. Non-SOR traders are constrained to trade on a single venue (typically the listing exchange) and therefore often trade at worse prices. The presence of SOR traders increases competition between liquidity providers since it increases the visible depth available to traders. This reduces the execution probability of limit orders on a single trading venue, and the execution probability of the lower-priority hidden orders relatively more compared to visible orders. The presence of more SOR traders should therefore reduce the volume share executed against hidden orders. However, more sophisticated SOR algorithms can also be used to detect the presence of hidden orders, as these offer an opportunity for depth improvement. Insofar our measure of SOR usage also captures the use of these more complex SORs, it is a priori unclear what effect is to be expected on hidden order trading.

For dark venues the relation between volume and the use of SORs is less ambiguous: there can only be volume in dark venues when traders access these venues. The more traders have access to dark venues, the larger the execution probability of dark orders and thus the more dark trading takes place. If our measure of SOR usage properly captures multi-market trading, thus including dark venues, we expect that SOR usage positively affects dark trading activity.

The fraction of traders that used SOR, $SOR_{i,t}$ is measured only on the consolidated lit market level.[15] We use the measure from van Kervel (2015), based on the fraction of trades that occurs simultaneously across markets, which are assumed to originate from SORs. A trade is taken to occur simultaneously with a previous trade on another venue when the best quotes of the executing venue have not changed, they are of the same sign and occur within 100 milliseconds. This leads to a dummy variable $S_{i,t,k}$ equal to one when a trade occurs simultaneously.[16] To account for the fact that a trade only occurs simultaneously when depth on one executing venue is not sufficient to fill the trade, the following linear regression is estimated for each stock i and day t on a trade-by-trade basis:

$$S_{i,t,k} = SOR_{i,t}P(x > T)_{i,t,k} + \epsilon_{i,t,k}$$

Where $P(x > T)$ denotes the probability that an order of size x exceeds the threshold T. Following van Kervel (2015) we set the threshold $T_{i,t,k}$ as the depth of the most liquid venue at time k and assume that order size is distributed exponentially with mean $\phi_{i,t}$, and thus $P(x > T)_{i,t,k} = exp(\frac{-T_{i,t,k}}{\phi_{i,t}})$. We estimate $\phi_{i,t}$ as the average trade size for stock i during day t.[17] [18] [19]

Table 3.4 shows that $SOR_{i,t}$ has a value of 7.43 percent on average, with a median of 7.03. These estimates are relatively low compared to the mean of 40 percent reported by van Kervel (2015), but his sample is characterized by a relatively larger fragmentation of trading across venues (39 percent is not traded on the listing exchange, compared to 22 percent on average in our sample, see Panel C of Table 3.1).

3.4.2 Methodology

We now turn to the question which market conditions affect the share of different types of opaque trading, *relative* to total trading activity. By investigating the determinants of the volume share of hidden order trading $\%HidV_{i,t}$ we gain insight in which circumstances (1) patient traders choose to submit a hidden or-

[15]Although the connectivity of traders to alternative trading venues may differ across venues.

[16]In practice, however, not all trades identified as simultaneous using this method originate from SORs, which can potentially lead to an overestimation of the fraction of traders using SOR. Simultaneous non-SOR trades are more likely on high volume days, hence we control for total volume in the regressions.

[17]We also use two alternatives to proxy for the fraction of SOR traders: (1) The daily average of the dummy $S_{i,t,k}$, and (2) The fraction of non-trade-throughs, similar to Foucault and Menkveld (2008). All measures are correlated and lead to qualitatively similar results.

[18]We only use information from simultaneous trades on *lit* trading venues. We assume that the fraction of SOR traders on lit venues is positively related to the fraction of SOR traders in total (i.e., the fraction of traders connected to multiple lit venues *and* dark venues).

[19]Although both our AT and SOR measures are proxies for algorithmic trading their correlation is quite low (near zero). This is in line with the assertion that both proxies measure fundamentally different styles of algorithmic trading. While AT is more likely to capture the behavior of HFT market makers (Hagströmer and Nordén, 2013), SOR is more related to optimal order execution.

der on a lit venue, and (2) impatient traders are more likely to trade against hidden depth. Similarly, we investigate the determinants of the volume share of dark trading $\%DarkV_{i,t}$. In addition to our primary definition of dark trading, the bulk of volume that is transacted in completely dark venues consists of block trades. Therefore we also examine the market conditions driving this volume component. Volume shares or relative volume measures are more suited for this analysis than absolute levels of trading as their scaling makes them independent of total trading activity. This makes relative volume a proxy for trading that relates more to the submission (or routing) decision and execution probability than to trading desire. We estimate the following model to asses how market and order book conditions impact different types of opaque trading.

$$\%OpaqueV_{i,t} \quad = \quad \boldsymbol{\gamma}'\mathbf{X}_{i,t,l} + \boldsymbol{\lambda}'\mathbf{Z}_{i,t,l} + \eta_{i,t,l} \tag{3.3}$$

$\%OpaqueV_{i,t}$ is either a measures for hidden order trading $\%HidV_{i,t,l}$, our measure of dark trading $\%DarkV_{i,t}$, or block trading $\%BlockV_{i,t}$. An intercept is not included because all variables in the model are standardized by stock and quarter. Market and order book characteristics, as defined in subsections 3.2.3 and 3.4.1, are contained in the vector $\mathbf{X}_{i,t,l}$ for the variables we assume to be endogenous ($Volume_{i,t,l}$, $VisDepth_{i,t,l}$, $QSpread_{i,t,l}$, $Volat_{i,t}$), and in the vector $\mathbf{Z}_{i,t,l}$ for the variables we believe are exogenous ($AT_{i,t,l}$ and $SOR_{i,t}$). Because of their potential endogeneity we instrument the variables $x_{i,t,l}$ by using their *market* version $x_{i'\neq i,t}$, as before. This approach is justified by the observation that, on the one hand, the variables included in our models are known to have a common market component to them. For instance, Chordia, Roll, and Subrahmanyam (2000) show there is a common market-wide component to liquidity while Hasbrouck and Seppi (2001) document commonality in liquidity, order imbalances, prices and volatility. On the other hand, it is hard to see how opaque trading activity in one stock could impact these market variables at the market level.

The full specification is:

$$\begin{aligned}\%OpaqueV_{i,t} \quad = \quad & \gamma_1 Volume_{i,t,Tot} + \gamma_2 VisDepth_{i,t,Lit} + \gamma_3 QSpread_{i,t,Lit} + \gamma_4 Volat_{i,t} \\ & +\lambda_1 AT_{i,t,Lit} + \lambda_2 SOR_{i,t} + \eta_{i,t} \end{aligned} \tag{3.4}$$

When $\%HidV_{i,t,Tot}$ and $\%DarkV_{i,t}$ are the dependent variables, they are measured relative to the total volume excluding block volume. Similarly, on the right hand side we have total volume $Volume_{i,t,Tot}$ with block volume excluded as a determinant, because block volume is an outlier-sensitive variable.

As an additional test, we estimate the determinants of hidden order trading on different subsamples based on trading venues. In these sepcifications we also distinguish between own venue depth and other venue depth to recognize that both may have an opposite impact on hidden order volume due to competition between trading venues. This leads to the following full specification of the model.

$$
\begin{aligned}
\%HidV_{i,t,l} &= \gamma_1 Volume_{i,t,l} + \gamma_2 VisDepth_{i,t,l} + \gamma_{2'} VisDepth_{i,t,l' \neq l} + \gamma_3 QSpread_{i,t,l} \\
&\quad + \gamma_4 Volat_{i,t} + \lambda_1 AT_{i,t,l} + \lambda_2 SOR_{i,t} + \eta_{i,t,l} \qquad (3.5)
\end{aligned}
$$

3.4.3 Results

Table 3.5 presents estimation results for different specifications of Equation (3.3). The results for $\%HidV$ are shown in columns (1 – 4), columns (5 – 8) present results with $\%DarkV$ as the dependent variable, while columns (9 – 12) show results for $\%BlockV$.

Total volume has a significant and strong positive effect on $\%HidV$ in all specifications. From column (1), if total volume increases with one standard deviation, the fraction of volume executed against hidden orders increases with 0.271 standard deviations. We interpret this as a positive effect from trading interest in a stock on hidden order trading volume. When the trading desire in a stock increases, and traders become more anxious to fill their trades, they are more likely to trade deeper in the limit order book and the execution probability of hidden orders increases. In contrast to hidden order trading, dark trading $\%DarkV$ is negatively affected by trading interest, as proxied by total volume. When total volume increases by one standard deviation, the fraction of volume executed dark decreases by 0.196 standard deviations. When the trading interest increases, so does the desire to trade immediately. Dark trading venues do not perform well in providing immediacy (Menkveld, Yueshen, and Zhu, 2015) and as a result, traders are more likely to switch to lit venues to execute their trades. Block trading activity is strongly positively affected by total volume, i.e. when there is a larger trading desire in the market, a larger portion of trading tends to go through block trading mechanisms. A one standard deviation increase in total volume leads to a 0.359 standard deviation increase in block volume. Given that block volume is about 27 percent of average daily market volume which represent only 7 trades of about 4.4 million euro on average this is hardly surprising. There are few alternatives to work trades of this size through the market other than through designated block trading mechanisms.

Volatility does not have a clear effect on hidden order and dark trading in our models. Only when volume is excluded from the specification does volatility become a significantly positive determinant for hidden order trading, and a negative determinant for dark trading, this due to the high correlation between volatility and volume measures. Indeed, increased volatility may reflect new information and induce investors to trade. The positive (negative) effect of volatility on hidden order (dark) trading activity is thus mainly indirect through volume. We find no evidence that an increase in volatility directly impacts opaque trading activity through an increase in hidden order submissions as predicted by Harris (1996) and Aitken, Berkman, and Mak (2001), or through a decrease in

Table 3.5: Determinants of Opaque Trading Volume

This Table presents estimation results of Equation (3.3) for different specifications. The dependent variable is either $\%HidV$ (the volume of hidden orders executed across all lit venues relative to the total consolidated volume), $\%DarkV$ (volume of non-block trades executed across all dark venues relative to the total consolidated volume) or $\%BlockV$ (the volume of block trades executed across all dark venues relative to the consolidated total volume). Independent variables in the models are: $Volume_{Tot}$, the consolidated volume across all venues, excluding block volume; $Volume_{Tot+Block}$, the consolidated volume across all venues, including block volume; $Volat$, the daily standard deviation of five-minute midquote returns; $VisDepth_{Lit}$, the time-averaged consolidated lit market depth quoted within 50 basis points of the midquote; $VisDepth_{l}$, the time-averaged lit venue depth quoted within 50 basis points of the consolidated midquote, based on Degryse, de Jong, and van Kervel (2015); $QSpread_{Lit}$, the time-averaged consolidated market quoted bid-ask spread (relative to the midquote); AT_{Lit}, a proxy for algorithmic trading based on Hendershott, Jones, and Menkveld (2011), and SOR, an estimate of the fraction of traders employing smart order routers, based on van Kervel (2015).

All equations are estimated by 2SLS on a panel of daily observations for 27 Dutch large cap stocks which spans 738 trading days, from November 2007 until September 2010. Variables are standardized by stock-quarter by subtracting the mean and dividing by the standard deviation. Endogenous independent variables $Volume_x$, $Volat$, $VisDepth_x$ and $QSpread_x$ are instrumented by using their 'market version', i.e. for each stock i and day t we take the cross-sectional mean of the variable excluding stock i. AT_x and SOR are assumed to be exogenous. Standard errors are Newey-West for panel data, t-statistics are shown in parentheses, and ***, ***, * indicates significance at the 1%, 5% and 10% level respectively.

	$\%HidV$				$\%DarkV$				$\%BlockV$			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Volume_{Tot}$	0.271***	0.302***		0.288***	-0.196***	-0.167***		-0.178***				
	(10.50)	(13.66)		(14.86)	(-7.42)	(-7.36)		(-9.20)				
$Volume_{Tot+Block}$									0.359***	0.328***		0.148***
									(12.29)	(12.94)		(6.27)
$Volat$	0.029	-0.032	0.244***		0.030	-0.003	-0.126***		-0.307***	-0.277***	-0.088***	
	(1.17)	(-1.37)	(12.62)		(1.22)	(-0.13)	(-6.78)		(-11.84)	(-11.17)	(-4.29)	
$VisDepth_{Lit}$	-0.118***	-0.140***	-0.065***	-0.124***	0.015	0.008	-0.023	0.009	-0.050**	-0.041*	0.022	0.019
	(-5.13)	(-6.19)	(-2.76)	(-5.43)	(0.71)	(0.39)	(-1.10)	(0.46)	(-2.26)	(-1.89)	(0.93)	(0.87)
$QSpread_{Lit}$	-0.144***	-0.151***	-0.188***	-0.132***	0.021	0.013	0.053**	0.033	0.025	0.031	-0.031	-0.116***
	(-5.92)	(-6.20)	(-7.48)	(-6.01)	(0.91)	(0.59)	(2.32)	(1.62)	(1.01)	(1.30)	(-1.23)	(-5.40)
AT_{Lit}	-0.131***		-0.246***	-0.126***	-0.075***		0.009	-0.069***	0.073***		-0.058***	0.023
	(-8.73)		(-25.03)	(-8.97)	(-5.16)		(0.93)	(-5.25)	(5.07)		(-5.72)	(1.53)
SOR	-0.078***		-0.066***	-0.078***	0.030***		0.022**	0.031***	-0.020**		-0.008	-0.023***
	(-9.24)		(-7.67)	(9.01)	(3.34)		(2.45)	(3.36)	(-2.50)		(-0.82)	(-2.64)
Obs	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416	17,416
R^2	0.123	0.102	0.076	0.125	-0.017	-0.016	0.000	-0.014	0.227	0.214	0.002	0.107

dark trading as predicted by Zhu (2014) and Menkveld, Yueshen, and Zhu (2015). Volatility has a strong and significant negative effect on block trading with a coefficient of -0.307 as traders tend to avoid block trades in volatile markets, either because the execution uncertainty is too large, or because the price uncertainty is large.

As expected, consolidated visible depth has a significant negative effect on hidden order trading, consistent over all specifications. If visible depth increases by one standard deviation, ceteris paribus, the fraction of volume executed against hidden orders decreases by 0.118 standard deviations (from column (1)). When more visible depth is quoted on lit venues this reduces the execution probability of hidden orders, thereby also reducing incentives to submit hidden orders. Our other measure of market quality, the consolidated quoted spread, has an opposite effect: hidden order trading is decreasing in the size of the spread (i.e., increasing in liquidity). When the spread increases with one standard deviation the fraction of trades executed against hidden orders reduces by 0.144 standard deviations. This is somewhat surprising, as theory predicts that patient large traders should be more likely to submit hidden orders when the spread is wider if hidden orders are used as a tool to limit competition from other liquidity suppliers (Buti and Rindi, 2013). However, the relation between the spread and hidden order usage may be more complex, as De Winne and D'Hondt (2007) also find a negative relation between spread size and the submission of hidden orders. Furthermore, the execution probability of hidden orders may also be hampered when the spread is wider, leading to more executed hidden orders when the spread is narrow.

Contrary to our expectations, lit market quality as measured by visible depth and the quoted spread does not significantly affect dark trading. One potential explanation is that market quality affects the venue submission choice for patient and impatient traders differently. A liquid lit market attracts order flow away from dark venues from impatient traders, but the increased competition for (hidden) order execution can crowd out patient traders such that they switch to another (dark) venue to submit their orders. When both effects offset each other the combined effect may be zero. For block trading, the coefficient for visible depth is negative and mildly significant in columns (9) and (10). Block trading is thus decreasing in lit venue liquidity, which suggests that order splitting in lit venues can be a substitute for block trading when lit venues are liquid. It is remarkable then that coefficient for visible depth is insignificant in the model for dark trading because non-block dark venues are more similar to lit venues than block venues.

Algorithmic trading consistently has a negative effect of around -0.130 on hidden order trading (when not controlled for total volume the effect even becomes more negative). There are three potential explanations. First, algorithms used by traders can substitute for hidden orders, reducing the amount of hidden orders submitted to lit trading venues. This is similar to the finding by De Winne and D'Hondt (2007) that 'principal' orders (which are submitted by brokers for their own account) are less likely to make use of hidden orders compared to 'client' orders (which are submitted by brokers on behalf of clients). The former are more sophisticated traders who tend to substitute hidden order usage by a better monitoring of the market. Similarly, algorithms may contain a fair amount of sophistication, but more importantly, they are designed to constantly monitor the market and trade whenever opportunities present themselves. Second, the presence of algorithmic traders increases competition for order flow among liquidity suppliers, especially when they also have a speed advantage, such as the case of HFTs. The increased competition then decreases the execution probability of those orders that have the lowest execution priority, i.e. hidden orders. Third, algorithmic trading is sometimes associated with concerns of predatory trading. If large patient traders (who would consider hidden orders) observe that the lit venues are more crowded with algorithms, they may refrain from trading here.

The presence of algorithmic traders also negatively affects dark trading when we control for volume. Two potential explanations also hold here. First, algorithms on lit venues may substitute for dark trading if they monitor the market closely and slice and dice large orders to the market. Second, insofar as algorithmic trading is perceived as toxic *and* when dark venues are perceived to be crowded with toxic algorithms, large traders may reduce their dark order submissions. The latter is consistent with the finding that, controlling for volume and volatility, algorithmic trading is positively related to block trading. When large traders fear order flow toxicity they may refrain from trading in any venue to which algorithms have access to, and instead resort to the designated block trading markets where information leakage is minimal.

The fraction of trades that is executed against hidden orders is also negatively affected by the usage of SORs by traders. When more liquidity takers use SOR to trade (simultaneously) on multiple lit venues hidden order trading declines. A one standard deviation in the fraction of SOR traders decreases the relative hidden order volume by 0.078 standard deviations. As with algorithmic trading there are two explanations, depending on *who* uses SOR. First, SOR can be used as a substitute by traders to execute their large orders through order splitting over multiple venues to find cheaper execution. Second, when more traders use SOR to tap into the liquidity offered at different trading venues the amount of visible depth with which hidden order traders are competing increases, reducing their execution probability. In turn this also reduces incentives to submit hidden orders.

The fraction of SOR traders is positively related to dark trading. When traders that are using SORs do not only connect to lit venues, but also engage in searches for liquidity in dark venues, an increase in the fraction of traders using the technology results in an increase in dark trading. By contrast, SOR usage is negatively related to block trading since SORs that split orders over multiple lit venues can substitute for block trading mechanisms.

In sum, hidden order trading and dark trading are differently impacted by most market and order book conditions, while both types of opaque trading are negatively impacted by algorithmic trading activity.

We now turn to the analysis of the determinants of hidden order trading across different trading venues. Panels A, B, C and D of Table 3.6 present estimation results of Equation (3.5) for four subsamples based on the executing venue. For the main trading venue (Panel A) results are highly similar to those with consolidated hidden order trading as the independent variable (Table 3.5) since hidden order volume on the main venue is on average 82 percent of total hidden order volume. The additional variable $VisDepth_{l' \neq l}$ has a positive and significant sign as expected. When visible depth on alternative venues increases with one standard deviation, the fraction of volume executed against hidden orders on the main venue increases by 0.079 standard deviations. This is in line with the explanation that as it becomes more difficult to obtain execution for hidden orders on the alternative venues, traders who wish to trade with hidden orders substitute those venues for the main venue.

Panels B, C and D show that volume is not a significant determinant of hidden order trading on alternative venues. Only on the largest alternative venue is volume significant in 3 out of 4 specifications where it is included, but not the full specification. An increase in trading interest on alternative venues (volume on venue l) does not increase hidden order trading on venue l. Own venue depth remains to have a significantly negative impact on hidden order trading, but other venue depth is generally insignificant. When visible depth is building up on competing lit venues hidden order traders do not substitute their venue for any of the alternative venues. A potential explanation is that hidden depth on competing lit venues constitutes for a large part of depth quoted on the main venue, and that the connectivity of traders to alternative venues is too limited to have a significant impact. The quoted spread, algorithmic trading and SOR have the same sign as before, and remain significant.

3.5 Conclusion

Financial markets offer different opportunities to trade opaquely. This chapter studies two forms of opaque trading – hidden order trading that can take place on several lit trading venues and dark trading, away from lit trading venues. Using a detailed high-frequency dataset our research provides insight in the segmentation of opaque trading into these two shades of opacity.

Our main results can be summarized as follows. First, using a simultaneous equations framework we show that dark trading and hidden order trading negatively affect each other. This supports the idea that dark trading and hid-

Table 3.6: Determinants of Hidden Order Volume Across Venues

This Table presents estimation results of Equation (3.5) for different trading venue subsamples. Panel A presents estimation results for different specifications for the Euronext subsample, Panel B for the MTF subsample, Panel C for the Chi-X subsample and Panel D for the Other MTF (Turquoise and Bats) subsample. The dependent variable is $\%HidV_l$ the volume of hidden orders executed across the lit venue(s) in the subsample relative to the total lit venue volume from the subsample. Independent variables in the models are: $Volume_l$, the lit venue volume; $Volat$, the daily standard deviation of five-minute midquote returns; $VisDepth_{Lit}$, the time-averaged consolidated lit market depth quoted within 50 basis points of the midquote; $VisDepth_l$, the time-averaged lit venue depth quoted within 50 basis points of the consolidated midquote, based on Degryse, de Jong, and van Kervel (2015); $QSpread_{Lit}$, the time-averaged consolidated market quoted bid-ask spread (relative to the midquote); AT_{Lit}, a proxy for algorithmic trading based on Hendershott, Jones, and Menkveld (2011), and SOR, an estimate of the fraction of traders employing smart order routers, based on van Kervel (2015).

All equations are estimated by 2SLS on a panel of daily observations for 27 Dutch large cap stocks which spans 738 trading days, from November 2007 until September 2010. Variables are standardized by stock-quarter by subtracting the mean and dividing by the standard deviation. Endogenous independent variables $Volume_x$, $Volat$, $VisDepth_x$ and $QSpread_x$ are instrumented by using their 'market version', i.e. for each stock i and day t we take the cross-sectional mean of the variable excluding stock i. AT_x and SOR are assumed to be exogenous. Standard errors are Newey-West for panel data, t-statistics are shown in parentheses, and ***, ***, * indicates significance at the 1%, 5% and 10% level respectively.

Panel A: Main Venue Hidden Order Trading

	(1)	(2)	(3)	(4)	(5)
$Volume_l$	0.275***	0.351***		0.285***	0.280***
	(8.64)	(14.91)		(11.00)	(8.65)
$Volat$	0.018	-0.045*	0.207***		-0.018
	(0.73)	(-1.90)	(9.71)		(-0.76)
$VisDepth_l$	-0.219***	-0.207***	-0.203***	-0.223***	-0.202***
	(-9.78)	(-9.21)	(-8.87)	(-10.08)	(-9.31)
$VisDepth_{l'\neq l}$	0.079***	0.021	0.136***	0.081***	0.102***
	(3.28)	(1.01)	(5.93)	(3.32)	(4.68)
$QSpread_{Lit}$	-0.068***	-0.102***	-0.064***	-0.060***	
	(-2.98)	(-4.83)	(-2.77)	(-2.73)	(0.00)
AT_{Lit}	-0.180***		-0.333***	-0.178***	-0.194***
	(-7.73)		(-26.90)	(-7.92)	(-8.95)
SOR	-0.054***		-0.060***	-0.054***	-0.054***
	(-6.42)		(-6.90)	(-6.40)	(-6.33)
Obs	15,564	15,564	15,564	15,564	15,564
R^2	0.139	0.116	0.101	0.139	0.141

Panel B: Alternative Venue Hidden Order Trading

	(1)	(2)	(3)	(4)	(5)
$Volume_l$	0.000	0.028		0.003	0.035
	(0.00)	(1.22)		(0.15)	(1.31)
$Volat$	0.004	-0.047**	0.004		-0.050**
	(0.18)	(-2.10)	(0.25)		(-2.31)
$VisDepth_l$	-0.070**	-0.154***	-0.070***	-0.070**	-0.013
	(-2.41)	(-5.89)	(-2.65)	(-2.43)	(-0.49)
$VisDepth_{l'\neq l}$	-0.031	0.070***	-0.031	-0.032	-0.009
	(-1.18)	(3.08)	(-1.25)	(-1.20)	(-0.33)
$QSpread_{Lit}$	-0.120***	-0.109***	-0.120***	-0.118***	
	(-5.21)	(-4.75)	(-5.25)	(-5.55)	
AT_{Lit}	-0.212***		-0.212***	-0.211***	-0.208***
	(-11.34)	(0.00)	(-16.52)	(-12.05)	(-11.21)
SOR	-0.107***		-0.107***	-0.107***	-0.100***
	(-10.10)	(0.00)	(-10.18)	(-10.22)	(-9.61)
Obs	15,564	15,564	15,564	15,564	15,564
R^2	0.050	0.012	0.050	0.050	0.050

Table 3.6 continued.

Panel C: Largest Alternative Venue Hidden Order Trading

	(1)	(2)	(3)	(4)	(5)
$Volume_l$	0.024	0.057***		0.034*	0.044*
	(0.97)	(2.65)		(1.76)	(1.78)
$Volat$	0.017	-0.044**	0.033*		0.001
	(0.72)	(-1.95)	(1.81)		(0.04)
$VisDepth_l$	-0.154*	-0.040	-0.156*	-0.175**	0.021
	(-1.86)	(-0.49)	(-1.89)	(-2.15)	(0.31)
$VisDepth_{l' \neq l}$	-0.093	0.036	-0.101	-0.110	0.050
	(-1.16)	(0.48)	(-1.28)	(-1.39)	(0.71)
$QSpread_{Lit}$	-0.080***	-0.070***	-0.084***	-0.078***	
	(-3.18)	(-2.82)	(-3.37)	(-3.15)	(0.00)
AT_{Lit}	-0.199***		-0.210***	-0.196***	-0.187***
	(-11.12)	(0.00)	(-17.25)	(-11.53)	(-10.48)
SOR	-0.071***		-0.068***	-0.071***	-0.066***
	(-7.31)	(0.00)	(-7.09)	(-7.39)	(-6.88)
Obs	15,564	15,564	15,564	15,564	15,564
R^2	0.040	0.005	0.041	0.040	0.042

Panel D: Other Alternative Venue Hidden Order Trading

	(1)	(2)	(3)	(4)	(5)
$Volume_l$	0.038	0.031		0.030	0.063*
	(1.09)	(1.19)		(1.05)	(1.84)
$Volat$	-0.014	-0.018	0.005		-0.049*
	(-0.46)	(-0.61)	(0.18)		(-1.72)
$VisDepth_l$	-0.202***	-0.193**	-0.203***	-0.192***	0.007
	(-2.65)	(-2.51)	(-2.66)	(-2.56)	(0.12)
$VisDepth_{l' \neq l}$	-0.080	-0.066	-0.093	-0.070	0.116*
	(-1.02)	(-0.85)	(-1.20)	(-0.91)	(1.78)
$QSpread_{Lit}$	-0.131***	-0.140***	-0.135***	-0.135***	
	(-4.14)	(-4.32)	(-4.35)	(-4.54)	
AT_{Lit}	-0.150***		-0.170***	-0.152***	-0.145***
	(-6.17)		(-11.87)	(-6.46)	(-5.91)
SOR	-0.144***		-0.139***	-0.143***	-0.149***
	(-11.83)		(-11.68)	(-12.02)	(-12.52)
Obs	11,465	11,465	11,465	11,465	11,465
R^2	0.051	0.020	0.050	0.050	0.050

den order trading, both forms of opaque trading, are substitutes. After deciding that a trader wants to reduce the visibility of his order in the market he decides whether he wants to trade in a dark venue, or whether he wants to trade with a hidden order in the lit market. This decision is likely to be based on personal preferences, order characteristics and the prevailing market conditions. However, we also find that the substitution is more likely from hidden orders to dark orders than the other way around. The effect from dark trading on hidden order trading is relatively small and insignificant. One explanation is that the relative difference in opacity between hidden orders and orders on dark venues makes the use of hidden orders a more inadequate substitute to dark trading than the other way around.

Second, we establish the determinants of hidden order trading and dark trading separately. Both are differently impacted by market and order book conditions. This is in line with the finding that both types of opaque trading are substitutes. We show that hidden order trading is increasing in total volume, while the opposite is true for dark trading. A larger trading desire tends to increase executions of hidden orders on lit venues, in particular on the listing exchange, while decreasing executions on dark venues. This is consistent with the conjecture that traders choose to trade on venues that offer more liquidity and immediacy when they have a stronger trading desire. Hidden order trading is decreasing in the size of visible depth offered, and is increasing as the spread narrows. Both measures of liquidity directly impact the execution probability of hidden orders. Contrary to the predictions of Buti, Rindi, and Werner (2015) dark trading is largely unaffected by market quality measures on lit venues, possibly because too few traders have access to dark venues. Smart order routers reduce the execution probability of hidden orders because they increase the competition from other trading venues. At the same time SORs may provide a substitute for large patient traders to execute their trades. Overall, hidden order trading is decreasing in the use of SORs, while dark trading is increasing, as more traders tap into dark liquidity. Algorithmic trading negatively affects both types of opaque trading, either because algorithms substitute for opaque orders, because they reduce execution probabilities due to a more fierce competition for liquidity supply, or because they increase the risk of predatory trading too much for large traders. The latter is consistent with the finding that block trading increases with algorithmic trading.

The large amount of hidden orders on lit venues and the growing market share of dark venues makes insight into opaque trading activity in all its appearances relevant for traders, market operators, brokers and regulators. From the regulatory perspective, the growing amount of dark trading has caused regulators to be concerned abouts its consequences for market quality and welfare. In particular, the European Commission's recent proposals on MiFID II aim to curb dark trading activity in two ways (European Commission, 2014). First, the

Commission aims to bring OTC volume to regulated trading venues by forcing brokers that match client orders to register either as an MTF or as a SI when trades are internalized. Second, on the regulated trading venues the Commission proposes to limit the use of the reference price waiver and negotiated trade waiver by imposing market share caps at 4 percent for an individual trading venues and 8 percent for the global market. Furthermore, the scope of the reference price waiver will be narrowed to midpoint matching only. The large in scale waiver remains unaltered in order to continue to allow the trading of very large blocks of shares without alerting the market. In addition, the order management facility waiver which allows for hidden order trading also remains in place without restrictions. Our findings suggest that if the proposed legislation comes into force without any further adjustments, traders who now make heavy use of dark trading venues, may have trouble in finding an suitable opaque alternative. Although dark trading can be a substitute for hidden order trading, hidden orders appear to be a less adequate substitute for dark venues. It is therefore questionable whether lit venues can offer truly opaque trading alternatives.

Concluding Remarks

This dissertation contributes to our understanding of the micro-level structure of financial markets. It contains three empirical studies that relate to the behavior of traders and how this affects market outcomes, in both a transparent and an opaque setting. While the first chapter is situated in a concentrated market, the second and third chapter explicitly take into account the effects of market fragmentation. The first and second chapter focus on information that is publicly available from transparent order books. By contrast, the third chapter specifically focuses on trading activity from opaque mechanisms. Results in each chapter are driven by *order choice*, either between different order types within a single venue, the venue selection itself, or a combination of both. We document how market outcomes are affected by order choice. On the one hand, order choice is both a determinant of short-term price effects and a driver of the segmentation of volume into different classifications. On the other hand, order choice is also affected by current market conditions, such as the state of the limit order book or the perceived liquidity across different trading venues.

Chapter 1 investigates why information from the public limit order book helps to predict short-term returns in order-driven markets. We find that the bulk of predictability can be explained by order choice considerations of uninformed traders. As traders predictably condition their order choice on the state of the book, or gradually switch to more aggressive orders when their previous orders remain unexecuted, they cause short-term price deviations. We find little evidence for the alternative explanation that relates predictability to informed trading in the limit order book. In particular, we find that predictability decreases with the time horizon – but it does not completely vanish. Especially for actively traded stocks predictability is the highest for shorter horizon returns. This suggests that the transitory (order-driven) component is more important than the permanent (information-driven) one. Furthermore, there is no clear intraday pattern that can tie predictability to informed trading. Finally, we show that, cross-sectionally, predictability is inversely related to informed trading, but positively to depth. Thus, return predictability is higher when there is less informed trading and the order book is more competitive. This is consistent with the bulk of return predictability being driven by order choice, and not informed trading.

The research on intraday return predictability is extended to a setting where multiple trading venues compete for order flow in Chapter 2. In particular, we address whether in a fragmented market, order book imbalances can predict returns on other trading venues, controlling for own venue order book imbalances. Order book imbalances are shown to be strong predictors of returns on another venue when they are associated with more competitive prices. We also document that prices from different trading venues tend to adjust to one another. Our results suggest that the listing exchange is the leading venue for trading activity and price discovery. However, the order book from an alternative venue can also contribute to return predictability, and it does so relatively more when its market share is larger. This is in line with traders adjusting their order choice to the state of all relevant order books.

Chapter 3 investigates two distinct ways in which traders can hide their trading intentions. One the one hand, they can choose to submit hidden orders on lit trading venues and hide within the visible limit order book. On the other hand, traders can also choose to submit orders to completely dark venues. We examine the relation between executed hidden order volume and dark volume, and find that dark trading and hidden order trading are substitutes. However, dark trading appears to be a better substitute for hidden order trading than the other way around. Furthermore, a number of market conditions differently affect hidden order trading and dark trading. In particular, hidden order trading is preferred over dark trading on high volume days, when the visible part of the order book shows more depth, the bid-ask spread is narrow and fewer traders employ smart order routers. Algorithmic trading negatively affects both types of opaque trading.

Our findings have implications for traders, market operators and regulators. As the amount of information that is now disseminated by trading venues and other data vendors is growing, financial markets are becoming ever-more transparent. Today's traders increasingly rely on this information as an input for their trading strategies. This holds for human traders, who try to keep track of information on multiple trading screens, but even more so for the myriad of algorithms that are shaping the way modern financial markets work. For instance, traders engaging in a market making strategy need to constantly adapt their best estimate of a security's fundamental value as they receive new information, as well as make forecasts about transient (liquidity-driven) price movements. Our results imply that information on the current state of the limit order book can be used by traders to predict short-term returns, and that these price effects are mostly transient. Optimally traders should use information obtained from *all* trading venues in order to improve their predictions. Not all information should be weighted equally though, as order book imbalances between the bid and ask side obtained from the quotes of more competitive trading venues strongly predict positive returns, while imbalances obtained from less competitive venues are weak predictors, or even predict reversals.

The advances in market automation and heavier reliance on algorithms are developments that go hand in hand with the increased market fragmentation. As trading activity becomes more dispersed across different mechanisms (limit order books, dark pools, broker/dealer crossing networks or internalizers), traders, especially institutions, need to exert more effort in the search for liquidity. Algorithms can play a key role in this search. But our results imply that not all algorithms are equal. For instance, we find that when traders rely more heavily on smart order routers, volumes on dark trading venues increase. Our proxy for algorithmic trading in general, which is based on the message-to-volume ratio and is likely to pick up HFT market making activity, negatively affects hidden order and dark trading, but positively affects block trading activity. The fragmentation of trading volume into different mechanisms, and the accompanying search for liquidity, remains to be one of the key challenges for today's traders.

The growing amount of market information that is disseminated, together with the increasing pace at which traders can react to it, also boosts the desire for opacity among large, institutional, traders. While previous research warrants caution for an over-reliance on dark trading venues in the market, due to potential negative effects on overall market quality, the fact remains that exposure costs in the visible market are too high for some traders. From the point of view of those traders, it is therefore positive that MiFID II will maintain the large in scale waiver (and even plans to expand it by reducing the thresholds), as well as the order management facility waiver, in order to facilitate some forms of opaque trading that are deemed less harmful for market quality. But at the same time MiFID II will introduce a double volume cap for trading on dark venues that do not fall under the scope of these waivers, i.e. those that fall under the reference price waiver and negotiated trade waiver. Volume will be capped at the 4 percent level for the individual trading venue, and 8 percent for the consolidated market. So in order to keep on trading opaquely, either traders will need to scale up their order size, or resort more to hidden orders. Our results imply that traders who currently rely on dark venues may have trouble in finding a suitable alternative, since hidden orders on lit trading venues are not adequate substitutes. However, most hidden orders in our sample are reserve orders. Other types of hidden orders, i.e. (midpoint) hidden pegged orders, are more similar to order types on dark venues, and thus could be better substitutes. In light of the regulatory changes to come, European market operators and investment firms should take it as a challenge to incorporate new mechanisms within lit trading venues that allow to bring back opaque volumes. This is the strategy for instance followed by the London Stock Exchange, as it announced to introduce midpoint hidden pegged orders that allow for a minimum execution size specification. The inception of new trading mechanisms to cater to the needs of large traders, however, goes beyond the simple dichotomy between hidden orders and dark venues. For instance, Bats Chi-X recently launched a lit periodic auctions book that operates alongside its existing continuous lit and dark books.

Bibliography

ADMATI, A. R., AND P. PFLEIDERER (1988): "A Theory of Intraday Patterns: Volume and Price Variability," *Review of Financial Studies*, 1(1), 3–40.

AITKEN, M. J., H. BERKMAN, AND D. MAK (2001): "The Use of Undisclosed Limit Orders on the Australian Stock Exchange," *Journal of Banking and Finance*, 25(8), 1589–1603.

AKTAS, N., E. DE BODT, F. DECLERCK, AND H. VAN OPPENS (2007): "The PIN Anomaly Around M&A Announcements," *Journal of Financial Markets*, 10(2), 169–191.

BEBER, A., AND C. CAGLIO (2005): "Order Submission Strategies and Information: Empirical Evidence from the NYSE," *Working Paper*.

BESSEMBINDER, H., M. PANAYIDES, AND K. VENKATARAMAN (2009): "Hidden Liquidity: An Analysis of Order Exposure Strategies in Electronic Stock Markets," *Journal of Financial Economics*, 94(3), 361–383.

BIAIS, B., P. HILLION, AND C. SPATT (1995): "An Empirical Analysis of the Limit Order Book and the Order Flow in the Paris Bourse," *Journal of Finance*, 50(5), 1655–1689.

BLOOMFIELD, R., M. O'HARA, AND G. SAAR (2005): "The "Make or Take" Decision in an Electronic Market: Evidence on the Evolution of Liquidity," *Journal of Financial Economics*, 75(1), 165–199.

——— (2015): "Hidden Liquidity: Some New Light on Dark Trading," *Journal of Finance*, 70(4), 2227–2274.

BOULATOV, A., AND T. J. GEORGE (2013): "Hidden and Displayed Liquidity in Securities Markets with Informed Liquidity Providers," *Review of Financial Studies*, 26(8), 2096–2137.

BROGAARD, J., T. HENDERSHOTT, AND R. RIORDAN (2014): "High-Frequency Trading and Price Discovery," *Review of Financial Studies*, 27(8), 2267–2306.

BROLLEY, M. (2015): “Should Dark Pools Improve Upon Visible Quotes? The Impact of Trade-at Rules,” *Working Paper*.

BRUGLER, J. A. (2015): “Into the Light: Dark Pool Trading and Intraday Market Quality on the Primary Exchange,” *Working Paper*.

BRUNNERMEIER, M. K., AND L. H. PEDERSEN (2005): “Predatory Trading,” *Journal of Finance*, 60(4), 1825–1863.

BUTI, S., AND B. RINDI (2013): “Undisclosed Orders and Optimal Submission Strategies in a Limit Order Market,” *Journal of Financial Economics*, 109(3), 797–812.

BUTI, S., B. RINDI, AND I. M. WERNER (2011): “Diving into Dark Pools,” *Working Paper*.

——— (2015): “Dark Pool Trading Strategies, Market Quality and Welfare,” *Working Paper*.

CAO, C., O. HANSCH, AND X. WANG (2008): “Order Placement Strategies in a Pure Limit Order Book Market,” *Journal of Financial Research*, 31(2), 113–140.

——— (2009): “The Information Content of an Open Limit-Order Book,” *Journal of Futures Markets*, 29(1), 16–41.

CESPA, G., AND T. FOUCAULT (2014): “Sale of Price Information by Exchanges: Does It Promote Price Discovery?,” *Management Science*, 60(1), 148–165.

CHAKRAVARTY, S., AND C. W. HOLDEN (1995): “An Integrated Model of Market and Limit Orders,” *Journal of Financial Intermediation*, 4(3), 213–241.

CHORDIA, T., R. ROLL, AND A. SUBRAHMANYAM (2000): “Commonality in Liquidity,” *Journal of Financial Economics*, 56(1), 3–28.

——— (2002): “Order Imbalance, Liquidity, and Market Returns,” *Journal of Financial Economics*, 65(1), 111–130.

CHORDIA, T., R. ROLL, AND A. SUBRAHMANYAM (2005): “Evidence on the Spead of Convergence to Market Efficiency,” *Journal of Financial Economics*, 76(2), 217–292.

——— (2008): “Liquidity and Market Efficiency,” *Journal of Financial Economics*, 87(2), 249–268.

CHUNG, D. Y., AND K. HRAZDIL (2012): “Speed of Convergence to Market Efficiency: The Role of ECNs,” *Journal of Empirical Finance*, 19(5), 702–720.

COMERTON-FORDE, C., AND T. J. PUTNIŅŠ (2015): "Dark Trading and Price Discovery," *Journal of Financial Economics*, 118(1), 70–92.

COPELAND, T. E., AND D. GALAI (1983): "Information Effects on the Bid-Ask Spread," *Journal of Finance*, 38(5), 1457–1469.

DANIËLS, T. R., J. DÖNGES, AND F. HEINEMANN (2013): "Crossing Network versus Dealer Market: Unique Equilibrium in the Allocation of Order Flow," *European Economic Review*, 62(0), 41–57.

DE WINNE, R., AND C. D'HONDT (2007): "Hide-and-Seek in the Market: Placing and Detecting Hidden Orders," *Review of Finance*, 11(4), 663–692.

DEGRYSE, H., F. DE JONG, AND V. VAN KERVEL (2015): "The Impact of Dark Trading and Visible Fragmentation on Market Quality," *Review of Finance*, 19(4), 1587–1622.

DEGRYSE, H., G. TOMBEUR, M. VAN ACHTER, AND G. WUYTS (2013): "Dark Trading," *Chapter 13 in Market Microstructure in Emerging and Development Markets, Robert Kolb Series in Finance (John Wiley & Sons, Inc.), Eds H. Kent Baker and Halil Kiymaz*.

DEGRYSE, H., M. VAN ACHTER, AND G. WUYTS (2009): "Dynamic Order Submission Strategies with Competition between a Dealer Market and a Crossing Network," *Journal of Financial Economics*, 91(3), 319–338.

DUARTE, J., AND L. YOUNG (2009): "Why Is PIN Priced?," *Journal of Financial Economics*, 91(2), 119–138.

DUONG, H. N., P. S. KALEV, AND C. KRISHNAMURTI (2009): "Order Aggressiveness of Institutional and Individual Investors," *Pacific-Basin Finance Journal*, 17(5), 533–546.

EASLEY, D., S. HVIDKJAER, AND M. O'HARA (2002): "Is Information Risk Determinant of Asset Returns?," *Journal of Finance*, 57(5), 2185–2221.

EASLEY, D., N. M. KIEFER, AND M. O'HARA (1996): "Cream-Skimming or Profit-Sharing? The Curious Role of Purchased Order Flow," *Journal of Finance*, 51(3), 811–833.

EASLEY, D., N. M. KIEFER, M. O'HARA, AND J. B. PAPERMAN (1996): "Liquidity, Information, and Infrequently Traded Stocks," *Journal of Finance*, 51(4), 1405–1436.

EASLEY, D., M. O'HARA, AND L. YANG (2013): "Differential Access to Price Information in Financial Markets," *Working Paper*.

EUROPEAN COMMISSION (2011): “Review of the Markets in Financial Instruments Directive (MiFID): Frequently Asked Questions,” *Press Release MEMO/11/716*.

——— (2014): “Markets in Financial Instruments Directive (MiFID II): Frequently Asked Questions,” *Press Release MEMO/14/305*.

FOLEY, S., K. MALINOVA, AND A. PARK (2013): “Dark Trading on Public Exchanges,” *Working Paper*.

FOLEY, S., AND T. J. PUTNIŅŠ (2015): “Should We Be Afraid of the Dark? Dark Trading and Market Quality,” *Working Paper*.

FOUCAULT, T. (1999): “Order Flow Composition and Trading Costs in a Dynamic Limit Order Market,” *Journal of Financial Markets*, 2(2), 99–134.

FOUCAULT, T., R. KOZHAN, AND W. W. THAM (2015): “Toxic Arbitrage,” *Working Paper*.

FOUCAULT, T., AND A. J. MENKVELD (2008): “Competition for Order Flow and Smart Order Routing Systems,” *Journal of Finance*, 63, 119–158.

FREY, S., AND P. SANDAS (2009): “The Impact of Iceberg Orders in Limit Order Books,” *Working Paper*.

GARVEY, R., T. HUANG, AND F. WU (2014): “Why Do Traders Choose Dark Markets?,” *Working Paper*.

GARVEY, R., AND F. WU (2009): “Intraday Time and Order Execution Quality Dimensions,” *Journal of Financial Markets*, 12(2), 203–228.

GLOSTEN, L. R. (1994): “Is the Electronic Open Limit Order Book Inevitable?,” *Journal of Finance*, 49(4), 1127–1161.

GLOSTEN, L. R., AND P. R. MILGROM (1985): “Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders,” *Journal of Financial Economics*, 14(1), 71–100.

GOETTLER, R. L., C. A. PARLOUR, AND U. RAJAN (2005): “Equilibrium in a Dynamic Limit Order Market,” *Journal of Finance*, 60(5), 2149–2192.

——— (2009): “Informed Traders and Limit Order Markets,” *Journal of Financial Economics*, 93(1), 67–87.

GOMBER, P., AND A. PIERRON (2010): “MiFID: Spirit and Reality of a European Financial Markets Directive,” *Working Paper*.

GOZLUKLU, A. (2014): “Pre-trade Transparency and Informed Trading: Experimental Evidence on Undisclosed Orders,” *Working Paper*.

GRESSE, C. (2006): “The Effect of Crossing-Network Trading on Dealer Market’s Bid-Ask Spreads,” *European Financial Management*, 12(2), 143–160.

——— (2015): “Effects of Lit and Dark Market Fragmentation on Liquidity,” *Working Paper*.

GRIFFITHS, M. D., B. F. SMITH, D. A. S. TURNBULL, AND R. W. WHITE (2000): “The Costs and Determinants of Order Aggressiveness,” *Journal of Financial Economics*, 56, 65–88.

HAGSTRÖMER, B., AND L. NORDÉN (2013): “The Diversity of High-Frequency Traders,” *Journal of Financial Markets*, 16(4), 741–770.

HANDA, P., R. A. SCHWARTZ, AND A. TIWARI (2003): “Quote Setting and Price Formation in an Order Driven Market,” *Journal of Financial Markets*, 6(4), 461–489.

HARRIS, L. (1996): “Does a Large Minimum Price Variation Encourage Order Exposure?,” *Working Paper*.

——— (1997): “Order Exposure and Parasitic Traders,” *Working Paper*.

HARRIS, L. E. (1998): “Optimal Dynamic Order Submission Strategies in Some Stylized Trading Problems,” *Journal of Financial Markets, Institutions and Instruments*, 7(2), 1–76.

HARRIS, L. E., AND V. PANCHAPAGESAN (2005): “The Information Content of the Limit Order Book: Evidence from NYSE Specialist Trading Decisions,” *Journal of Financial Markets*, 8(1), 25–67.

HASBROUCK, J., AND G. SAAR (2009): “Technology and Liquidity Provision: The Blurring of Traditional Definitions,” *Journal of Financial Markets*, 12(2), 143–172.

——— (2013): “Low-latency trading,” *Journal of Financial Markets*, 16(4), 646–679.

HASBROUCK, J., AND D. J. SEPPI (2001): “Common Factors in Prices, Order Flows, and Liquidity,” *Journal of Financial Economics*, 59(3), 383–411.

HATHEWAY, F., A. KWAN, AND H. ZHENG (2014): “An Empirical Analysis of Market Segmentation on US Equities Markets,” *Working Paper*.

HAUTSCH, N., AND R. HUANG (2012): “On the Dark Side of the Market: Identifying and Analyzing Hidden Order Placements,” *Working Paper*.

HENDERSHOTT, T., C. M. JONES, AND A. J. MENKVELD (2011): “Does Algorithmic Trading Improve Liquidity?,” *Journal of Finance*, 66(1), 1–33.

HENDERSHOTT, T., AND H. MENDELSON (2000): "Crossing Networks and Dealer Markets: Competition and Performance," *Journal of Finance*, 55(5), 2071–2115.

IRVINE, P. J., G. J. BENSTON, AND E. KANDEL (2000): "Liquidity Beyond the Inside Spread: Measuring and Using Information in the Limit Order Book," *Working Paper*.

JAIN, P. K. (2005): "Financial Market Design and the Equity Premium: Electronic versus Floor Trading," *Journal of Finance*, 60(6), 2955–2985.

JAIN, P. K., P. JAIN, AND T. H. MCINISH (2011): "The Predictive Power of Limit Order Book for Future Volatility, Trade Price, and Speed of Trading," *Working Paper*.

KANIEL, R., AND H. LIU (2006): "So What Orders Do Informed Traders Use?," *Journal of Business*, 79(4), 1867–1913.

KUMAR, P., AND D. J. SEPPI (1994): "Limit and Market Orders with Optimizing Traders," *Working Paper*.

KWAN, A., R. MASULIS, AND T. H. MCINISH (2015): "Trading Rules, Competition for Order Flow and Market Fragmentation," *Journal of Financial Economics*, 115(2), 330–348.

MENKVELD, A. J. (2013): "High Frequency Trading and the New Market Makers," *Journal of Financial Markets*, 16(4), 712–740.

MENKVELD, A. J., B. Z. YUESHEN, AND H. ZHU (2015): "Shades of Darkness: A Pecking Order of Trading Venues," *Working Paper*.

MOINAS, S. (2010): "Hidden Limit Orders and Liquidity in Order Driven Markets," *Working Paper*.

NAES, R., AND B. A. ODEGAARD (2006): "Equity Trading by Institutional Investors. To Cross or Not to Cross?," *Journal of Financial Markets*, 9(2), 79–99.

NAES, R., AND J. A. SKJELTORP (2006): "Order Book Characteristics and the Volume-Volatility Relation: Empirical Evidence from a Limit Order Market," *Journal of Financial Markets*, 9, 408–432.

NIMALENDRAN, M., AND S. RAY (2014): "Informational linkages between dark and lit trading venues," *Journal of Financial Markets*, 17, 230–261.

PAGANO, M. (1989): "Trading Volume and Asset Liquidity," *Quarterly Journal of Economics*, 104(2), 255–274.

PARDO, A., AND R. PASCUAL (2012): "On the Hidden Side of Liquidity," *European Journal of Finance*, 18(10), 949–967.

PARLOUR, C. A. (1998): “Price Dynamics in Limit Order Markets,” *Review of Financial Studies*, 11(4), 789–816.

PASCUAL, R., AND D. VEREDAS (2009a): “Does the Open Limit Order Book Matter in Explaining Informational Volatility?,” *Journal of Financial Econometrics*, pp. 1–31.

——— (2009b): “What Pieces of the Limit-Order Book Information Matter in Explaining Order Choice By Patient and Impatient Traders,” *Quantitative Finance*, 9(5), 527–545.

RANALDO, A. (2004): “Order Aggressiveness in Limit Order Book Markets,” *Journal of Financial Markets*, 7(1), 53–74.

READY, M. J. (2014): “Determinants of Volume in Dark Pool Crossing Networks,” *Working Paper*.

ROCK, K. (1996): “The Specialist’s Order Book and Price Anomalies,” *Review of Financial Studies*, 9, 1–20.

ROŞU, I. (2015): “Liquidity and Information in Order Driven Markets,” *Working Paper*.

ROSENBLATT SECURITIES (2012): “Let There Be Light: Rosenblatt’s Monthly Dark Liquidity Tracker,” .

——— (2013): “Let There Be Light: Rosenblatt’s Monthly Dark Liquidity Tracker - European Edition,” .

SECURITIES AND EXCHANGE COMMISSION (2010): “Concept Release on Equity Market Structure,” *Release No 34-61358, File S7-02-10*.

SEPPI, D. J. (1997): “Liquidity Provision with Limit Orders and a Strategic Specialist,” *Review of Financial Studies*, 10(1), 103–150.

THEISSEN, E., AND G. WUYTS (2010): “Empirical Tests of Models of Dynamic Limit Order Markets,” *Working Paper*.

TOMBEUR, G., AND G. WUYTS (2015): “Limit Order Book Information and Return Predictability,” *Working Paper*.

TUTTLE, L. (2006): “Hidden orders, trading costs and information,” *Working Paper*.

VAN KERVEL, V. (2015): “Competition for Order Flow with Fast and Slow Traders,” *Review of Financial Studies*, 28(7), 2094–2127.

WEAVER, D. G. (2014): “The Trade-At Rule, Internalization, and Market Quality,” *Working Paper*.

WHITE, M. J. (2014): "Enhancing Our Equity Market Structure," *Speech (Available online: http://www.sec.gov/News/Speech/Detail/Speech/1370542004312#.VOMZ1S4YFnA)*.

YAMAMOTO, R. (2012): "Intraday Technical Analysis of Individual Stocks on the Tokyo Stock Exchange," *Journal of Banking & Finance*, 36(11), 3033–3047.

YAO, C. (2012): "Hidden Agendas: A Study of the Impact of Concealed Orders," *Working Paper*.

YE, M. (2011): "A Glimpse into the Dark: Price Formation, Transaction Cost and Market Share of the Crossing Network," *Working Paper*.

ZHU, H. (2014): "Do Dark Pools Harm Price Discovery?," *Review of Financial Studies*, 27(3), 747–789.

Doctoral dissertations from the Faculty of Business and Economics, see:
http://www.kuleuven.ac.be/doctoraatsverdediging/archief.htm.